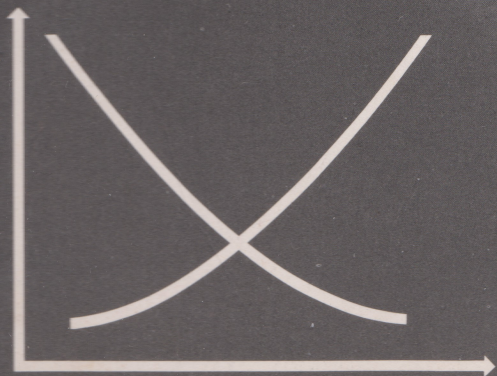




Economy and Society





THE OPEN UNIVERSITY

UNDERSTANDING SOCIETY: A FOUNDATION COURSE UNITS 10-13

Economy and Society

THE OPEN UNIVERSITY PRESS

CONTENTS

UNIT 10 ECONOMIC WANTS

F. S. Brooman

Page 7

11 PRODUCTION AND SUPPLY

F. S. Brooman

Page 31

12 MARKETS AND PRICES

L. Wagner

Page 57

13 THE SOCIOLOGY OF ECONOMIC BEHAVIOUR

C. J. Tunstall

Page 87

Open University courses provide a method of study for independent learners through an integrated teaching system, including textual material, radio and television programmes and short residential courses. This text is one of a series that make up the correspondence element of the Social Science Foundation Course.

The Open University's courses represent a new system of university-level education. Much of the teaching material is still in a developmental stage. Courses and course materials are, therefore, kept continually under revision. It is intended to issue regular up-dating notes as and when the need arises, and new editions will be brought out when necessary.

Further information on Open University courses may be obtained from The Admissions Office, The Open University, P.O. Box 48, Bletchley, Buckinghamshire.

The Open University Press
Walton Hall Bletchley Bucks

First published 1971 Reprinted 1971
Copyright © 1971 The Open University

All rights reserved
No part of this work may be
reproduced in any form, by
mimeograph or any other means,
without permission in writing
from the publishers

Designed by The Media Development Group of The Open University

Printed in Great Britain by

EYRE AND SPOTTISWOODE LIMITED
AT GROSVENOR PRESS, PORTSMOUTH

SBN 335 01502 6

Open University courses provide a method of study for independent learners through an integrated teaching system, including textual material, radio and television programmes and short residential courses. This text is one of a series that make up the correspondence element of the Social Sciences Foundation Course.

The Open University's courses represent a new system of university-level education. Much of the teaching material is still in a developmental stage. Courses and course materials are, therefore, kept continually under revision. It is intended to issue regular up-dating notes as and when the need arises, and new editions will be brought out when necessary.

Further information on Open University courses may be obtained from The Admissions Office, The Open University, P.O. Box 48, Bletchley, Buckinghamshire.

INTRODUCTION TO UNITS 10-13

At this point, the course moves from Social Psychology to Economics. Having described the way people are 'socialised' into members of the community, it now turns to the question of their livelihood. This should not, however, be seen as a sharp break in the subject-matter of the course, because the formative processes described in the last four units have important effects on the working of the economy, and are themselves conditioned in many ways by it. A simple agricultural society, for example, needs fewer skills than a complex manufacturing economy, and they are mostly of the kind that are better learned in the field than in the classroom; so its educational system is likely to be less elaborate than that of an industrial society, and with a much less clear distinction between work, play and formal instruction. We shall be looking in more detail at certain socio-economic relationships in Unit 13.

Back in Unit 3, the *economic problem* of society in general was defined as that of allocating scarce resources between competing uses.¹ For a single individual, it is a question of deciding how far to satisfy each of his various wants, given that his means of doing so (namely, the income he can obtain from his resources) are not enough to satisfy all his wants simultaneously. But exactly what his wants are, will depend to a considerable extent on the kind of socialisation process he has passed through; thus in any kind of society he will want food, but whether it is to be wheat or rice, roast beef or whale-blubber, depends on the way his tastes have been formed by growing up in one kind of social context rather than another. In some societies, the individual's consumption-pattern is actually prescribed in some detail by custom or even by law, which specifies (for example) the kind of clothes he may wear and the sort of house he may live in; in these cases, the individual's freedom of choice is considerably reduced, and his personal 'economic problem' is correspondingly simplified. Custom may limit choice on the production side, too – for example, where sons must follow the occupations of their fathers, or where workers practise time-honoured crafts in which innovations are taboo; similarly, choice may be heavily influenced by a social 'value system'² in which (say) manual occupations are despised in comparison with non-manual ones, or military virtues are prized above those which make for economic success. For a whole society to be willing to accept such constraints on individual choice, it is necessary that its members should have passed through an appropriate conditioning-process in childhood and adolescence, and without some knowledge of the process it will not be possible to understand why the economy works as it does. Again, for further discussion of these and related questions see Unit 13.

1 See Unit 3, p. 64.

2 A social 'value' is something people generally consider good, such as peace, or a high standard of living, or truthfulness. But some good things may be thought still better than others: a 'value system' is the order of priorities among them.

In Units 10–12 inclusive, however, it will be assumed that there are no important institutional constraints on individuals' freedom of choice between different goods and services. Whatever the set of wants and preferences they have formed through the socialisation process, it will be supposed that it is not so rigid as to call for fixed quantities of all commodities under all circumstances. In order to exhibit the logic of economic choice in a completely general way, it is assumed to operate in a money-using 'market' economy of the kind described in Unit 3. Each individual has certain resources – his labour-power, and property of various kinds accumulated in the past – from which he derives a money-income that he can spend quite freely on the goods and services which are on sale in the various markets that make up the economy. In Unit 10, attention will be focussed on the way wants are translated into market demand, in terms of the offer of money for goods and services; the next will show how the scarcity of resources affects production, and in the following one (Unit 12) the working of markets will be more closely examined. It is hoped that the student will bear in mind throughout that the assumptions just mentioned are far from being universally valid, and may require severe modification in the case of particular societies (Unit 13).

SUPPLEMENTARY MATERIALS (bound separately)

Self-assessment tests

Computer marked assignments

Tutor marked assignments

Radio notes

Television notes

Notes on the correspondence material

THE SET BOOKS

Understanding Society. * Readings in the Social Sciences edited by The Social Sciences Foundation Course Team. London, The Open University Press/Macmillan (1970).

Introduction to Economic Analysis by J. M. Fleming. London, Allen and Unwin (1969).

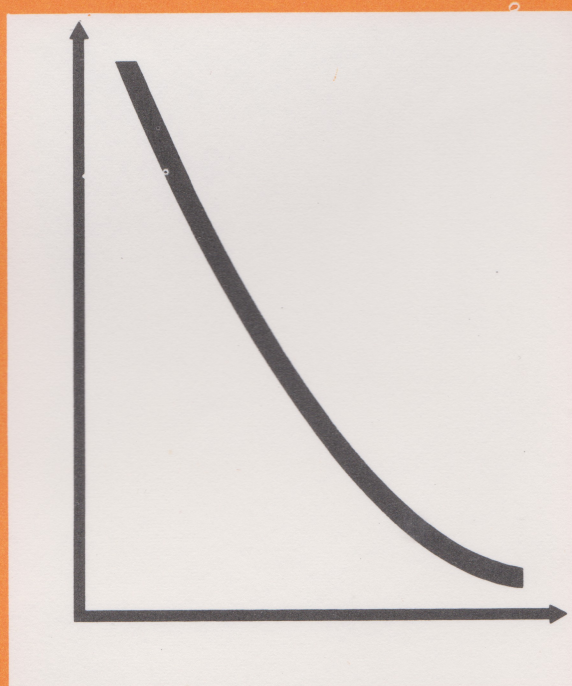
Introducing Sociology by Peter Worsley. London, Penguin Books (1970).

Society by E. Chinoy. New York, Random House (1967).

* Referred to in the text as the *Reader*.

Unit 10

Economic Wants



CONTENTS UNIT 10

	PAGE
1 THE CONCEPT OF MARKET DEMAND	11
2 THE CONCEPT OF UTILITY	14
3 THE PRINCIPLE OF EQUI-MARGINAL RETURNS	17
4 THE THEORY OF CONSUMER BEHAVIOUR	20
5 INDIVIDUAL AND MARKET DEMAND CURVES	23
6 THE ELASTICITY OF DEMAND	24
7 SUMMARY AND CONCLUSIONS	26
APPENDIX	
1. Note on the Measurement of Utility	27
2. Mathematical Notes	29
BIBLIOGRAPHY	30

ECONOMIC WANTS

I THE CONCEPT OF MARKET DEMAND

In Unit 3, individuals were described as allocating their spending between different commodities according to (a) the intensity of their wants for the commodities, and (b) the prices at which the commodities were available. This meant that, in the market for any given commodity, the amounts of it which the whole body of consumers were willing to buy would vary with the price, in some such way as that illustrated in the imaginary example given in the following table (reproduced from Unit 3):

Price (pence per unit)	30	25	20	15	10
Quantity demanded (thousand units per day)	100	120	150	190	250

The same kind of information can be shown in a two-dimensional diagram like Figure 10.1 below, in which the horizontal scale (OQ) measures quantities of the commodity in thousands of units per day, and the vertical scale OP measures its price in pence; both measurements begin from the left-hand bottom corner where the scales meet (the 'origin' of the diagram, as it is called).

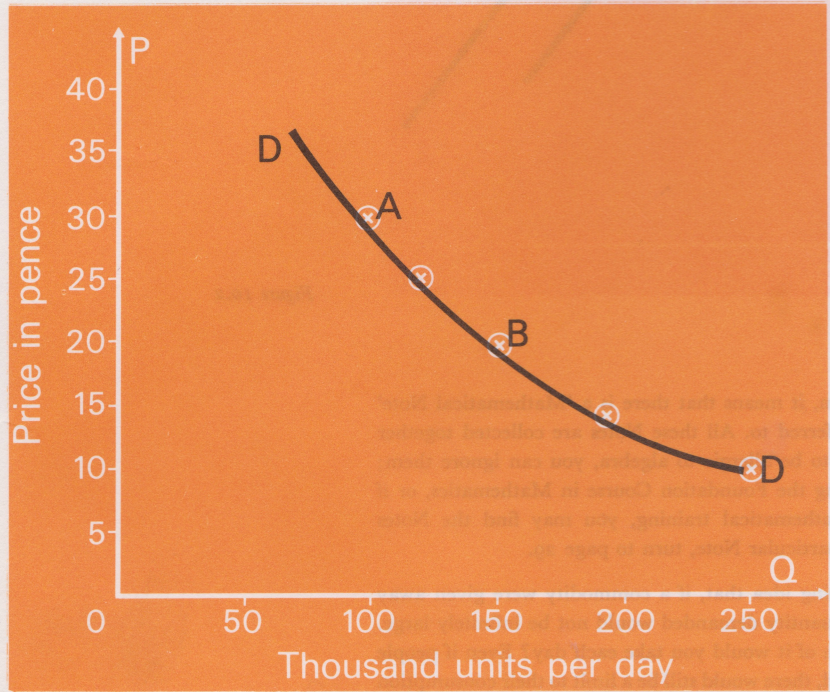


Figure 10.1

The crosses indicate the quantities associated with the various prices in the table – for example, 'A' shows that 100,000 units are wanted at a price of 30p, 'B' that 150,000 units are wanted at 20p, and so on. The curve DD – the *demand curve* – passes through all these points, and indicates the general relationship

between the price of the commodity and the quantity demanded.¹ The shape of DD suggests that if it were prolonged to the left, it would cut the vertical scale at some point; this means that there must be some price which is so high that no one would think it worthwhile to buy the commodity at that price. Similarly, prolonging DD to the right would make it cut the horizontal scale; this would indicate the maximum amount of the commodity that would be wanted, even if it were given away free.²

Two points must be emphasised with regard to the interpretation of any demand curve. The first is that it must have a 'time-dimension' if it relates to the flow of a commodity from sellers to buyers;³ it is necessary to specify whether purchases are 'per week', 'per day', 'per year' or per any other suitable time-period. Obviously, the number of eggs a housewife will want to buy per week, at a given price, will be smaller than the number she will want to buy in a month and larger than the number she will want to buy in a single day. The second point is that the whole curve indicates a set of

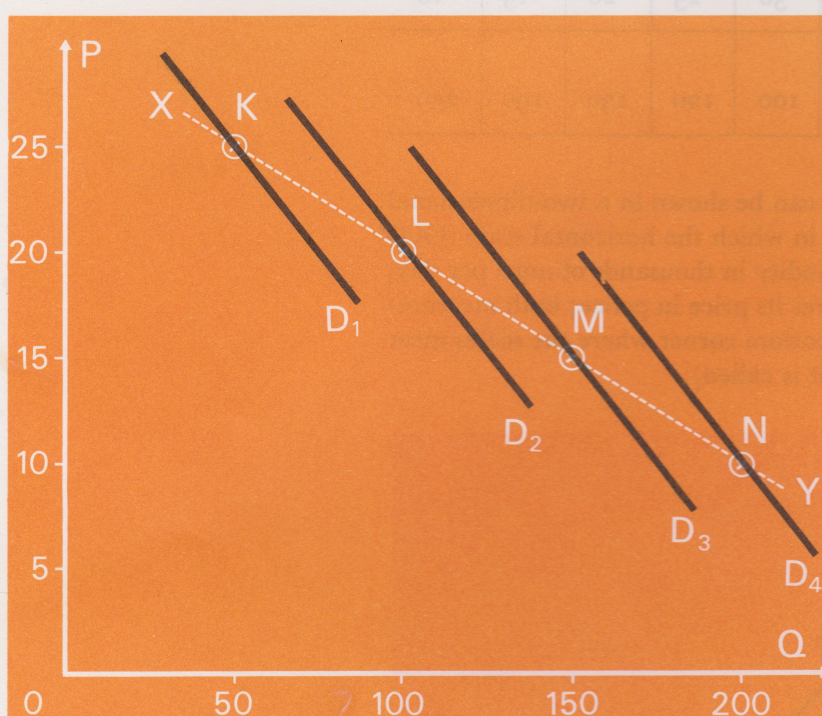


Figure 10.2

M 1. Whenever you see this sign, it means that there is a 'Mathematical Note' on the part of the text referred to. All these Notes are collected together at the end of the Unit. If you happen to be allergic to algebra, you can ignore them; on the other hand, if you are also taking the Foundation Course in Mathematics, or if you happen already to have some mathematical training, you may find the Notes helpful. So if you want to consult this particular Note, turn to page 29.

2 Students sometimes find it a surprising idea that, if a commodity were given away free of charge (i.e. at zero price), the quantity demanded would not be infinitely large. But suppose bread were free: how much of it would you take each day? Even if people were to substitute it for other kinds of food, there would still be a limit to their consumption of it; beyond that limit, it would actually be a nuisance to have more bread.

3 It is, however, possible to define demand in terms of the quantity of a commodity that people want to keep in hand as a *stock* at a given moment – for example, the number of tins of beans a housewife keeps in her larder. The time-dimension disappears in this case; there is no 'flow' from sellers to buyers, because there is no time for a flow to take place. It is sometimes convenient to speak of a 'stock demand' as opposed to a 'flow demand', but for the present all demands are to be understood as 'flows'.

alternative possibilities within the same time-period, rather than a sequence of actual purchases during successive periods. It is not possible to build up a demand curve by seeing how much of a commodity people buy when the price is 25p, then seeing how much they buy when the price is 20p, doing the same when the price is 15p and 10p, and then drawing a curve through the corresponding points on a graph (K, L, M and N in Fig. 10.2). The fact that only one price can actually be in force at any given moment means that there must inevitably be a time-interval between any two observations; during the intervals, consumers' preferences may change so as to shift the whole demand curve bodily from position D_1 to D_2 , and again from D_2 to D_3 , and yet again from D_3 to D_4 . Each of the observations K, L, M and N will then relate to a *different* demand curve, and the broken line XY will not itself be a demand curve in the sense defined above.¹

The nature of the price-quantity relationship can vary considerably from one commodity to another; for example, if the curve DD in Figure 10.1 were a steeply-sloping one, this would indicate that buyers were not very sensitive to price-changes, and would neither be stimulated by price-reductions to increase their purchases very much, nor impelled by price-increases to make large cuts in the quantities they buy. A gently-sloping demand curve, by contrast, would mean that quite small changes in price would cause large variations in the quantity demanded – as, for example, where a close substitute for a commodity is available and a small increase in the latter's price causes buyers to turn to the substitute in large numbers. It can happen, too, that the demand curves for some commodities vary in slope, being steep over a certain price-range and gently inclined elsewhere. But whatever the individual variations in the shapes of particular commodities' demand curves, they all have one general feature in common – namely, that they slope downwards from left to right, meaning that other things being equal,² the quantity demanded of any commodity varies inversely with its price. This may at first seem to be a mere matter of common sense; most people, no doubt, would agree that they tend to buy more of a good when its price comes down and less when its price goes up. All the same, it is worth going more deeply into the reasons for this, because they provide further insights into the logic of

1 Because only one market price can actually exist at a given moment, and only one quantity of the commodity can be changing hands, it might seem that it is not possible to verify more than one point on any particular demand-curve. However, the remainder of the curve could be filled in by questioning representative samples of consumers as to how much they would be buying if the price were different from its present level – by market research, in fact.

2 This is a phrase much used by economists, who often give it in its Latin form *ceteris paribus* (or *cet. par.* for short). In the present context, it means that whatever may happen to the price of the commodity under consideration, no other commodity's price is allowed to alter; the output of everything else is supposed to remain unchanged; so are the incomes of all producers and consumers, whose pattern of likes and dislikes is also assumed fixed. In fact, the only things that *can* change, once the 'other things being equal' formula has been uttered, are those directly under analysis – in this case, the price of the commodity and the quantity demanded by buyers. These are the 'variables' of the analysis; all other things are 'constants' – or rather, since they could be imagined as changing in other circumstances, 'temporary constants' or 'parameters'.

economic behaviour. For this purpose, it is necessary to consider in more detail the motivation of the individual consumer.¹

TEST 1

THE CONCEPT OF UTILITY

The typical consumer's basic situation has already been briefly sketched on p. 6. At any time, he possesses certain resources from which he can obtain an income. He can hire out his labour-power to someone else in return for wages, or use it in an enterprise of his own from which (if all goes well) he receives a profit. His accumulated wealth – the result of his own saving in the past, of inheritance, and/or of windfalls – may already be in a form, such as government bonds, company shares or 'real' property, which yields income in money; it may be in a form which provides income 'in kind', such as the accommodation given by an owner-occupied dwelling-house; or it may be held in money, which itself returns no income but which can be used to acquire income-yielding assets. Just how much income he gets from his labour-power and property² depends on the level of wages he can earn, the rate of yield on his assets and the total amount of them. He may also receive a pension, social security benefits, or regular gifts from others. Given his total income from all sources, he can spend it on any of the goods and services available in the market in whatever way he thinks fit; he can also leave some of it unspent, or he can overspend his income by getting into debt or running down his assets. Both income and spending must be given a time-dimension; they are *flows* whose amounts will differ according to whether they are measured per day, per month or per year.

Suppose, then, that the individual decides in the light of his total income and assets to spend a certain amount of money per week on consumers' goods and services. There is a vast array of commodities to choose from, and a very large number of alternative ways of distributing his spending among them; his aim is to find the combination of purchases which satisfies his wants to a greater extent than any other. Exactly what commodities he chooses will obviously depend on the pattern of his tastes. There will be certain things, no doubt, which he will never buy under any circumstances because his dislike for them is absolute – tobacco, for example, in the case of a lifelong non-smoker; other things which he will always buy because they satisfy very intense and urgent wants;³ and still others which, though not regarded by him as essential requirements, are likely to be included among his purchases because they give him some

1 For the test and the answer to it, see the self-assessment questions in the NOTES AND EXERCISES.

2 Some economists, notably the 'Chicago school' associated with Professor Milton Friedman, speak of 'human and non-human wealth', the former including all forms of productive brain- and muscle-power, the latter all other assets. Most economists, however, normally define 'wealth' as accumulated property, i.e. 'non-human' assets in the Chicago sense. It should be noted that, on either of these definitions, 'wealth' is a *stock* of possessions existing at a moment of time, and should not be confused with income and output, which are *flows* over a period of time.

3 You may find it helpful at this point to refer back to the discussion of needs and wants on pp. 63–64 of Unit 3.

degree of satisfaction. However, it would be incorrect to conclude that he will distribute his expenditure in proportion to the total satisfaction he expects to obtain from each of the various commodities, even though such a suggestion might seem plausible at first sight. Consider the example of food: this is clearly a vital necessity, which will be put ahead of everything else in order of priority, because it satisfies wants which are greater and more urgent than all others. Yet the fact is that the average consumer does not spend most of his income on food;¹ in fact, the larger his income, the smaller the proportion of it allotted to food expenditure.² The reason is that once enough food has been bought to provide the consumer's normal diet, the desire to have still more becomes much weaker than his desire for the basic minimum had been; to have additional food may still increase his total satisfaction, but it will do so by less and less as his intake increases. At some point, therefore, he will find that spending a further 10p on food gives him less satisfaction than he could get by spending it on something else, and he will switch his expenditure to other commodities. If this point is reached fairly quickly, the consumer will end up spending a relatively small proportion of his income on food, even though the *total* satisfaction obtained from it is greater than that from all other commodities.

The principle involved here is known as the 'Law of Diminishing Marginal Utility'. 'Utility' is merely another word for the satisfaction of wants, in the phrase originally used in Unit 3, it is the 'well-being' derived from a given commodity. 'Marginal utility' is the *extra* satisfaction obtained from having an *additional* unit of the commodity; and the 'Law' is the proposition that the greater the total amount of the commodity consumed by an individual per period of time, the less will be the increase in total satisfaction gained by adding one more unit to his consumption. This is exemplified in the following numerical example. Suppose there is some means of measuring³ the degree of utility derived by a particular individual from different quantities of a given commodity, and that when measurements are made the results are as follows:

TABLE 1

Units of the commodity consumed per week	1	2	3	4	5	6	7	8
Total degrees of utility	10	19	27	34	40	45	48	49

1 In 1969, out of total expenditure of £28,590 millions, British consumers spent only £5,977 millions, or 21 per cent, on food. (Figures from *Economic Trends*, H.M.S.O.).

2 One of the oldest-established empirical generalisations in economics is 'Engel's Law'. Ernst Engel (not to be confused with Marx's friend Friedrich Engels) investigated the income-expenditure patterns of middle and working class families in Saxony in the 1850s, and found that the higher the level of income, the smaller was the proportion spent on food. All subsequent inquiries have found a similar inverse relationship.

3 The question of measuring utility is dealt with in a separate Note at the end of the main text (pp. 27-28); this is to avoid interrupting the argument at this point.

These are alternative possibilities which must be thought of as co-existing within the same time-period;¹ that is, if the consumer were to be given (say) four units of the commodity per week he would derive 34 degrees of utility from them, whereas if he were given (say) seven units per week instead, he would experience 48 degrees of utility. *Marginal* utility, at any given level of consumption, is the total utility derived from that number of units *minus* the total utility of the immediately preceding number; for example, if the 34 degrees derived from four units are subtracted from the 40 degrees derived from five, the difference – 6 degrees – is the additional, or *marginal*, utility which the consumer gets by having five units of the commodity per week instead of four.² In the same way, the marginal utility of seven units is 48 (the total utility of seven) *minus* 45 (the total utility of six units), i.e. 3 degrees. When similar calculations are made for all levels of consumption, the results are as shown in Table 2.

TABLE 2

Units per week	1	2	3	4	5	6	7	8
Degrees of marginal utility	10	9	8	7	6	5	3	1

The figures in the table clearly exemplify the Law of Diminishing Marginal Utility, since they show marginal utility to be smaller, the higher the level of weekly consumption; this inevitably follows from the sequence of figures in Table 1, which shows that although *total* utility increases as the number of units consumed per week rises, it does so *at a diminishing rate*.³ However, an illustrative example is not a proof: the ‘Law’ is worth considering only if it conforms to the facts, and since you – the student who is reading this – must yourself be a consumer, you may care to pause at this point and ask yourself whether it is borne out by your own experience. If, for example, you were to double the amount of food you eat per day, would you get twice as much enjoyment from it? If you had ten times as many clothes to wear as you have now, would you derive

1 In other words, the table does not show the result of giving the consumer one unit in a given week, two units the next week, three the one after that, and so on; the measurements are not *successive*. If they were, the possibility would arise that the consumer’s likes and dislikes might have changed between one week and another. Some commodities are ‘habit-forming’, so that appetite grows with what it feeds on; the consumer might then need (say) six units a week to give him the same satisfaction as that which had been given by a single unit some time earlier.

2 It might seem simpler to speak of ‘the marginal utility of the fifth unit’. But to think in terms of *successive* instalments, as ‘first’, ‘second’, ‘third’, etc. tempt one to do, is to risk the confusions mentioned in the previous footnote. To say ‘a second cup of tea gives me less enjoyment than the first, and a third cup less than the second’ could be quite untrue if I am allowed time to become thirsty again between the successive cups. Also, to speak of ‘the marginal utility of the fifth unit’ suggests that it is the utility given by this particular unit which is being referred to. But it could just as well be defined as ‘the utility which would be lost if any one of my present five units were taken away’ – in which case, no particular unit could be identified as being the ‘fifth’ one.

ten times as much satisfaction from them? The crux of the matter is the *satiability* of economic wants. If you can imagine yourself having so much of a commodity that you would have no use for any more of it, you are envisaging a situation in which total utility has reached a maximum. When total utility can no longer increase, marginal utility must be zero. If, then, the marginal utility of any commodity must eventually become zero because all wants are satiable, it is reasonable to suppose that it will gradually diminish towards that level as the quantity consumed is increased – so that the sequence of figures shown in Table 2 can be regarded as typical.

TEST 2

3 THE PRINCIPLE OF EQUI-MARGINAL RETURNS

The 'Law of Diminishing Marginal Utility' will now be assumed to hold good for the purposes of the argument which follows. The object is to explain how the individual consumer will allocate his expenditure between the various commodities which are available to be bought, given the pattern of his likes and dislikes, and given also that he has a fixed amount of money to spend in total.

The answer is that, if he wishes to maximise the total utility derived from all commodities, he must buy them in quantities such that the last (or 'marginal') 5p spent on any single commodity brings him the same enjoyment (or 'utility') as the last 5p spent on any other commodity; in other words, the marginal utility of 5p worth of anything he buys must be equal to that of 5p worth of anything else.¹ A simple numerical example will show the necessity for this. Suppose the consumer is choosing between two commodities, A and B. Table 3 shows the utility to him of various quantities of the two commodities; but instead of measuring quantities in the usual physical units (such as ounces of tobacco and yards of cloth), the table counts them in 'shillingworths' (where a 'shilling' is defined as 5p).² It is assumed that the utility derived from A is not in any way affected by the amount of B which is being consumed, and *vice versa*, and that the prices of A and B are constant throughout. Now suppose that the consumer has four shillings to spend on the two commodities. If he had to spend all four on one or the other of them (i.e., if no 'mixtures' were allowed) he would obviously buy B, since he would obtain 116 degrees of utility, as compared with the 68 degrees he would get from four shillingworth

¹ This assumes that it is not possible to buy less than 5p worth of anything at a time. If, on the contrary, most goods were actually available in pennyworths, the rule would be to distribute expenditure so that the last penny spent on any one good yielded the same satisfaction as the last penny spent on anything else. But there are some things, such as cars and houses, whose minimum unit value is much higher than 5p, even when they are put on a daily or weekly basis by, e.g., calculating the appropriate hire-purchase repayment rate per day or per week. If the minimum weekly outlay on buying a car happened to be £5, the consumer would have to ensure that it was not possible to buy a collection of other goods and services, worth £5 in total, which would give him greater satisfaction than the use of his car for a week. This is a less simple principle than the one given above.

² Although this unit will reach you after decimalisation has begun – hence the use of '5p' – it is assumed that everyone will still be referring to the 5p coin as a 'shilling' for some time to come, and that the word can be used as an equivalent for 5p without causing confusion.

TABLE 3

Total Utility	No. of Shillingworths (5p each)
A	
20	1
38	2
54	3
68	4
80	5
90	6
98	7
104	8
108	9

TABLE 3

No. of shillingsworths per day	Total Utility (degrees)		Marginal Utility (degrees)	
	A	B	A	B
1	20	50	20	50
2	38	80	18	30
3	54	100	16	20
4	68	116	14	16
5	80	128	12	12
6	90	136	10	8
7	98	140	8	4
8	104	143	6	3
9	108	145	4	2

of A. But if he can choose a *mixture* of the two commodities, it will pay him to substitute a shillingworth of A for one of the four of B: by reducing his purchase of B from four shillingworth to three, he reduces total utility from 116 degrees to 100 (i.e. he gives up a marginal utility of 16 degrees), but by buying one shillingworth of A he obtains 20 degrees, giving him a net gain of 4 degrees and making the utility of his total expenditure 120. However, it would not pay him to go further and substitute a second shillingworth of A for one of his three shillingworths of B, since he would then be losing a marginal utility of 20 degrees in order to obtain one of only 18 degrees. The best way of laying out his four shillings, then, is to spend three on B and one on A; this way, total utility is a maximum, and the marginal utilities of the commodities are equal.

If he had ten shillings to spend instead of four, the figures in Table 3 indicate that he should spend five of them on A and five on B, since the marginal utilities are then equal at 12 degrees; the total utility obtained from his ten shillings will then be 208 degrees. If he were to switch one of his shillings from A to B, he would gain 8 degrees of utility from the additional shillingworth of B, but would lose 12 by giving up one of A, so that total utility would fall to 204 degrees. Similarly, if he switched a shilling from B to A, he would gain 10 degrees but lose 12, ending up with a total of 206 degrees. So the pair of expenditures at which the marginal utilities of the commodities are equal is that which yields the greatest total utility. In just the same way, it can be shown that an expenditure of sixteen shillings should be divided so that nine are spent on A and seven on B; while a total expenditure of thirteen shillings should be made up of seven shillingworth of A and six of B.¹

The general principle for maximising satisfaction, then, is that expenditure should be allocated in such a way that the marginal utility of a shillingworth of any commodity is equal to that of any other. Although only two commodities are mentioned in the

1 The table shows some levels of total spending at which it is impossible to reach equal marginal utilities: for example, the best way of spending 14 shillings is to spend eight on A and six on B, even though their marginal utilities are still not equal in this situation. They are, however, less unequal than they would be if the expenditure were divided in any other way; if it were possible to buy less than a shillingworth of A and B, marginal utilities could be equated by spending rather less than eight shillings (say, 39p) on A and rather more than six (say, 31p) on B.

numerical example just given, it will apply in exactly the same way where there are three or more. This can be written in symbols as

$$MU_A = MU_B = MU_C = \dots = MU_N \quad (1)$$

where MU stands for marginal utility, and the letters which are tagged on at the foot (the 'subscripts' as they are called) stand for shillingsworths of goods A, B, C, etc.; the series of dots between MU_C and MU_N stand for the MU's of other commodities (D, E, F, etc.); the total number of commodities is denoted by N, and MU_N is the marginal utility of a shillingsworth of the Nth commodity which ends the series.

Alternatively, the principle can be stated in terms of physical units of the various commodities instead of shillingsworths of them. Suppose commodity 'A' is priced at a penny a unit, so that a shillingsworth is five units; then the MU of a physical unit of 'A' will be a fifth of the MU of a shillingsworth. To convert MU_A (the MU of a shillingsworth) into MU_a (the MU of a physical unit) it is necessary to multiply MU_A by the price of A in shillings – in this case, by one-fifth (or 0.2), since 1p is a fifth of a shilling. In symbols, $MU_A \times P_a = MU_a$, where P_a is A's price in shillings; if both sides of this equation are divided by P_a , it becomes

$$MU_A = \frac{MU_a}{P_a}$$

and if this is substituted into equation (1), along with the equivalent expressions for MU_B , MU_C , etc., the result is

$$\frac{MU_a}{P_a} = \frac{MU_b}{P_b} = \frac{MU_c}{P_c} = \dots = \frac{MU_n}{P_n} \quad (2)$$

Equation (2) is merely a way of saying that when the consumer is getting the highest possible degree of satisfaction from his total expenditure, he must be buying the various commodities in quantities such that the proportion between marginal utility and price is the same for all of them. Thus, if P_b is 5p and P_a is 1p, MU_b must be five times as large as MU_a if equation (2) is to be satisfied.

The advantage of equation (2) is that it specifies the prices of all commodities, instead of leaving them implicit as in equation (1).

But both equations express the same general principle – namely, that the consumer is best off when the last unit of expenditure on any one commodity brings him the same return in terms of utility as the last unit spent on any other, so that he is getting 'equi-marginal returns' (or 'equal returns at the margin'). This is an important principle in economics, with many applications apart from the present one; for example, a business firm selecting inputs of materials, labour, machinery, etc., for a manufacturing process will find that it pays to seek 'equi-marginal returns' in terms of the output accruing from its expenditures on each of the various inputs. It is even possible to speak of 'equi-marginal returns from time', which would be achieved if one allotted one's time so that it was not possible to make a more profitable use of it by (say) devoting five more minutes to reading at the expense of spending five minutes less on some other activity.

TEST 3

4 THE THEORY OF CONSUMER BEHAVIOUR

The original point of introducing the 'equi-marginal principle', however, was to explain how a consumer should allocate his spending between commodities if he wishes to maximise the total satisfaction obtained from it. With its aid, it is now possible to be more specific about the factors which account for the consumer's demand for any particular commodity.

(a) Basic to everything else are his *tastes* – i.e. the set of likes and dislikes which together determine the utility he derives from any given amount of any commodity. If his income, wealth and other circumstances (such as family responsibilities) are assumed fixed, and if the prices of commodities A and B are also constant, the figures in Table 3 will be entirely a reflection of his tastes.

(b) Given these tastes, and with the prices of commodities still assumed constant, the amount he wishes to spend on all consumable commodities will depend on his *income*. The larger this is, the more he will buy of everything¹ and the further down Table 3's columns he will be able to go. The more he buys, the lower the values of MU_A , MU_B , etc., in equation (1), because of the Law of Diminishing Marginal Utility. Total consumption spending need not, of course, be exactly equal to income; the consumer may underspend (i.e. save) or overspend ('dis-save'). How much he can overspend depends mainly on his *wealth*, i.e. his total assets (money, shares, 'real' property, etc.) *minus* his debts; and the greater his wealth, the less the incentive to underspend in order to accumulate still more of it. So his total consumption will be larger, the greater his wealth as well as the higher his income.

(c) Now suppose that his tastes, income and wealth are all assumed to be fixed, but that prices are free to change. If *all* prices were to change simultaneously in the same proportion, for example all rising by 10 per cent, the effects would be much the same as if his income had fallen by 10 per cent in money terms instead. If, on the other hand, some prices rise and others fall in such a way as to leave the overall average unchanged, the general purchasing-power of his income will not be affected, and whatever alterations he makes in his pattern of spending will be due entirely to the change in prices *relatively* to one another. In practice, of course, a change in any particular commodity's price is likely to set up a mixture of both these effects. If P_A (the price of good 'A') rises, while P_B , P_C , etc., stay the same as before, the purchasing-power of the consumer's income will be reduced – not very much if A happens to be only a minor item in his budget like salt or sealing-wax, but by an appreciable amount if A is an important commodity like meat. To the extent that this fall in 'real' income (i.e. in the income's purchasing-power) causes the consumer to buy less A, the rise in P_A will have exerted what is known as an *income effect*. But the fact that A has now become more expensive relatively to other goods means, in addition,

1 Except where he feels he can now afford to buy higher-quality goods in place of the cheaper substitutes he had been consuming before his income rose; he will now be buying *less* of these 'inferior goods'. In this sense, potatoes may be 'inferior' to beefsteaks: when his income was lower, he could not afford so many expensive steaks, and instead bought more potatoes to supplement his diet; so when his income rises, he will buy fewer potatoes and eat more steaks.

that the consumer is likely to substitute other goods for A in his allocation of spending, and to the extent that he reduces his buying of A for this reason, the rise in P_a will have had a substitution effect. In Table 3, a rise in P_a means that the purchaser gets a smaller quantity of A for each shilling, so that the 'total utility' figures in the first column will all be reduced; this will also cause the figures in A's 'marginal utility' column to fall,¹ and he will be led to rearrange his spending, buying less A than before. In terms of equation (2), a rise in P_a will cause $\frac{MU_a}{P_a}$ to fall; to restore its equality with $\frac{MU_b}{P_b}$, $\frac{MU_c}{P_c}$, etc., the consumer must buy less A, so that (in accordance with the Law of Diminishing Marginal Utility) MU_a will rise sufficiently to offset the rise in P_a .

(d) Thus, a rise in P_a will reduce the quantity demanded of good A (and, of course, a *fall* in P_a will *increase* it). But it is also likely to affect the demand for other goods as well. It is possible that a given rise in P_a might cause the demand for A to fall by the exact percentage needed to leave the consumer spending precisely the same amount of money on it as he did before² – in which case, his expenditure on other commodities will be unchanged. But if his total expenditure on A actually falls (because the quantity demanded is reduced in a greater proportion than the rise in P_a) he will have more money left to spend on other goods and his demands for B, C, etc., will increase; while if his total expenditure on A rises, his demand for other goods will fall. In the same way, a change in the price of B or C could cause the demand for A itself to alter, even though P_a itself had not changed.

(e) So far, it has been assumed that all commodities are in some sense substitutes for one another, so that if the price of A rises, the consumer will cut down his expenditure on it and buy B or C instead. There is an obvious sense in which (say) butter and margarine, or tea and coffee, are substitutes for one another; but even goods so apparently dissimilar as petrol and beer are substitutes to some extent, since they both compete for the consumer's expenditure. Some commodities, however, are so far from being substitutes that they actually go together as complements, like tea and milk, or bacon and eggs, or motor-cars and petrol. If an increase in the price of cars reduces the demand for them, it will cause the demand for petrol to fall as well; the effects are exactly the opposite of those which occur as between a pair of substitutes.³

1 E.g., if P_a doubles, all A's figures, both total and marginal, will be reduced by half.

2 Suppose P_a is 10p to begin with, and that the quantity of A demanded at this price is 10 units per day; total expenditure is 100p. If, when P_a rises to 12p, the quantity demanded falls to $8\frac{1}{3}$ units per day, total expenditure will still be exactly 100p.

3 Where goods are 'complementary', the assumption made earlier on page 17 breaks down; it is no longer reasonable to suppose that the utility of a given quantity of A is not affected by the amount of B being consumed, if B is complementary to A. To have more bread will give less satisfaction, if there is no butter to put on it, than it will if more butter is provided at the same time; one's enjoyment of a car is conditional on how much petrol one has in the tank. To allow for this, Table 3 would have to be made much more complicated, showing a different 'total utility' column under 'A' for each separate quantity of B, and *vice versa*.

To sum up: the quantity of any commodity A demanded by the consumer per period of time depends on the following: the price of the commodity itself (P_a); the prices of other commodities ($P_b, P_c, \dots, P_n$); the consumer's income (Y); his wealth (W); and his tastes (T). In symbols,

$$D_a = D(P_a, P_b, P_c, \dots, P_n, Y, W, T) \quad (3)$$

where D_a is the quantity of A demanded, and the ' $D(\dots)$ ' means 'depends on'.¹ If all the magnitudes represented inside the bracket were held constant except P_a (that is, if other prices, income, wealth and tastes are treated as 'parameters' or 'temporary constants'), the relationship would be as shown in Figure 10.3 below:² the higher the level of P_a , the smaller the amount of D_a ; the lower P_a , the greater D_a . Curve $D_a(1)$ shows the precise relationship which holds when certain initial values of the parameters are assumed (e.g. that the consumer's income is £20 a week and his assets £100; that the price of B is 10p a pint, that of C 15p a pound, and so on; and that his tastes are as described by an appropriately extended version of Table 3). If any of the parameters were to change, the curve would shift to right or left; for example, a rise in income would be likely to push it out from $D_a(1)$ to a new position $D_a(3)$, showing that the consumer would want to buy more³ A even though P_a had not changed at all; while if P_b (the price of a substitute commodity) were to fall, the curve would move to $D_a(2)$.

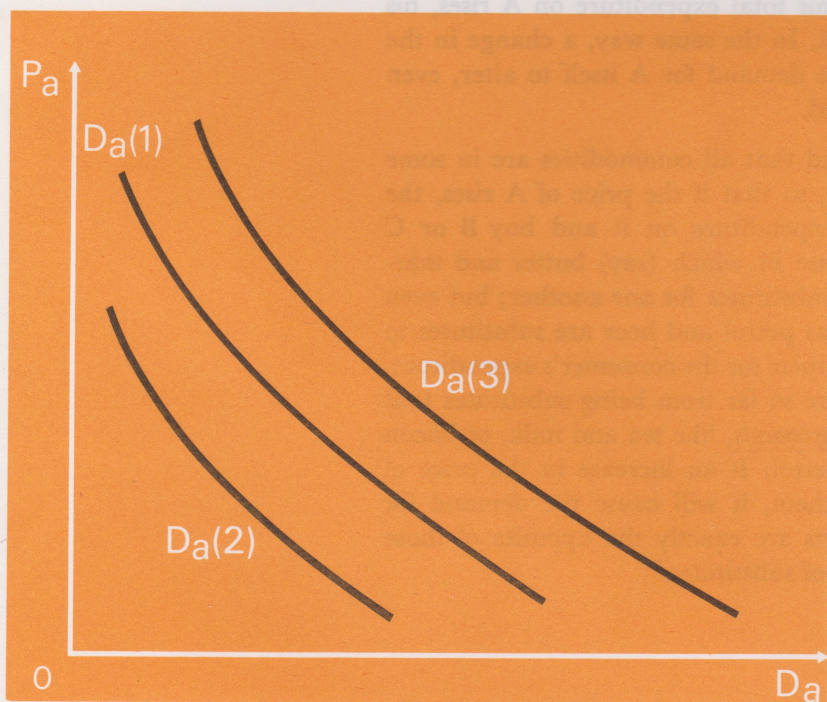


Figure 10.3

1 In mathematical language ' $D(\dots)$ ' could be read as 'is a function of'. The symbol indicates not only that there is a relationship between the two sides of the equation, but also that it is a regular, systematic one which could be defined more precisely.

2 No figures are marked on the scales of this diagram, because no precise quantities are referred to in the text. But, as in previous diagrams, the price increases from O upwards on the vertical scale, and the quantity demanded increases rightwards from O on the horizontal scale.

3 Unless A happens to be an 'inferior' good of the kind referred to in footnote 1, page 20, in which case a rise in income will cause the consumer to want less of it.

5 INDIVIDUAL AND MARKET DEMAND CURVES

The D_a curves of Figure 10.3 bear an obvious resemblance to the 'demand curves' of Figures 10.1 and 10.2 – the difference being that while the former relate to a single individual, the latter describe the propensities of the whole mass of consumers in the entire market. The 'market' demand curve is simply the horizontal sum of all individual ones. Suppose, for example, that there are three consumers, Smith, Brown and Jones, whose individual demand curves for commodity A are those drawn in Figures 10.4a, 10.4b and 10.4c. When they are combined (or 'aggregated') into a market demand curve by adding up the quantities demanded by Smith, Brown and Jones at each possible price,¹ the result is as shown in Figure 10.4d.

Figure 10.4a, b, c

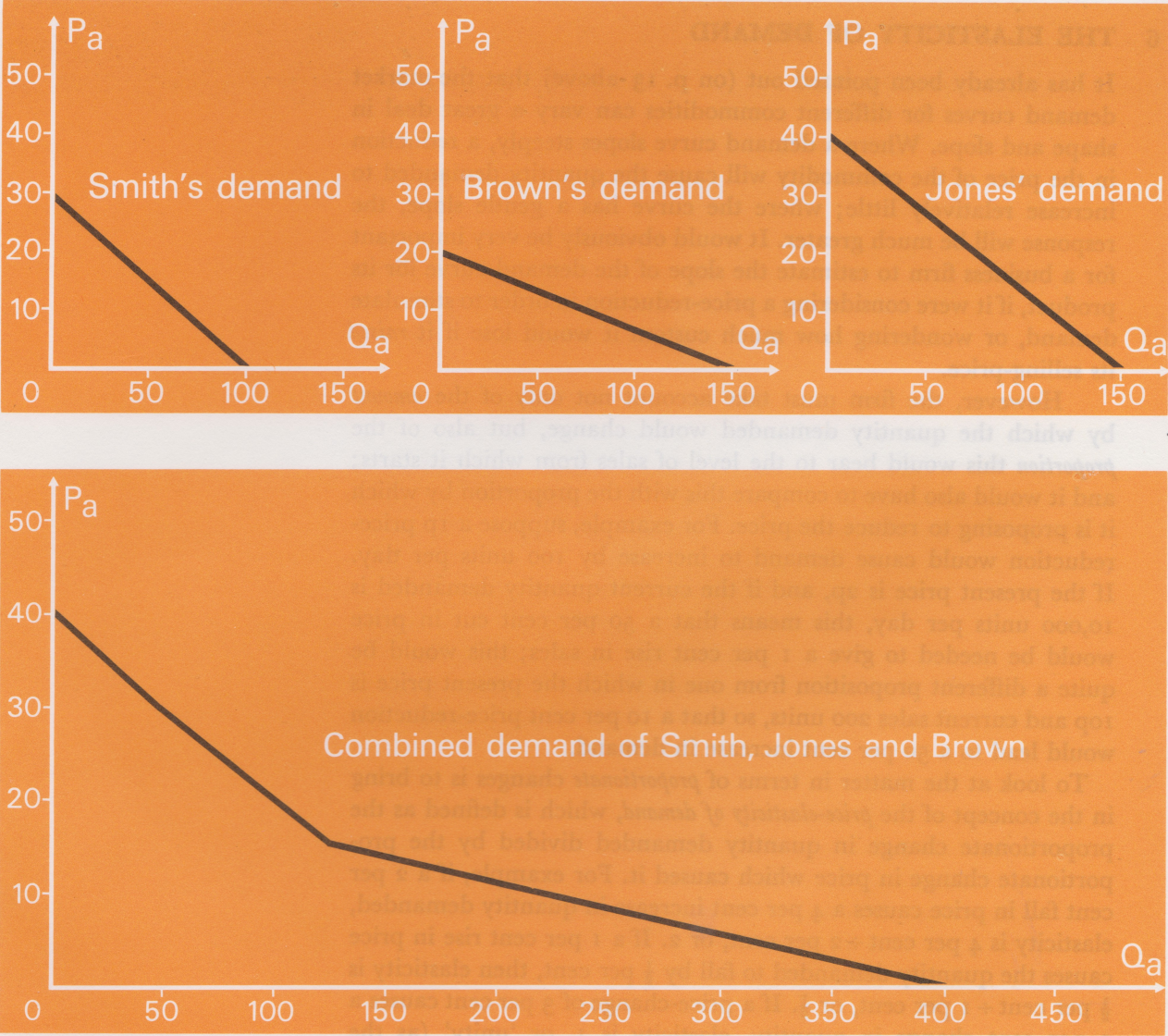


Figure 10.4d

¹ For example, when the price of A is 20p, Smith wants to buy about 30 units, Brown does not want to buy any units at all, and Jones wants to buy 70 units; so in the 'combined demand' diagram (Fig. 10.4d) the three of them together are shown as wanting to buy 100 units at a price of 20p. At prices above 30p, only Jones wants to buy, so the 'combined demand curve' in Fig. 10.4d is identical with Jones' curve in Fig. 10.4c between 40p and 30p. At prices between 30p and 20p, Smith wants to buy as well as Jones, so the 'combined' curve sums up the demands of the two of them. At prices below 20p, Brown wants to buy as well, so the 'combined' curve is the sum of all three individual demands at or below 20p.

The parameters of this market curve are a combination of those of the individuals' curves – the prices of other commodities, the incomes and wealth of the individuals, and their tastes. In most markets for consumers' goods, the number of buyers is very large, so that a change in a single person's tastes or income will have only a very small effect on demand in the market as a whole; it is when such changes are widespread (including changes in the distribution of income and wealth between individuals, as well as general increases or decreases in them) that they are likely to cause perceptible shifts in the market demand curve.

TEST 4

6 THE ELASTICITY OF DEMAND

It has already been pointed out (on p. 13 above) that the market demand curves for different commodities can vary a great deal in shape and slope. Where a demand curve slopes steeply, a reduction in the price of the commodity will cause the quantity demanded to increase relatively little; where the curve has a gentle slope, the response will be much greater. It would obviously be very important for a business firm to estimate the slope of the demand curve for its product, if it were considering a price-reduction in order to stimulate demand, or wondering how much custom it would lose if it raised its selling-price.

However, the firm must take account not only of the *amount* by which the quantity demanded would change, but also of the *proportion* this would bear to the level of sales from which it starts; and it would also have to compare this with the proportion by which it is proposing to reduce the price. For example, suppose a 1p price-reduction would cause demand to increase by 100 units per day. If the present price is 2p, and if the current quantity demanded is 10,000 units per day, this means that a 50 per cent cut in price would be needed to give a 1 per cent rise in sales; this would be quite a different proposition from one in which the present price is 10p and current sales 200 units, so that a 10 per cent price-reduction would lead to a 50 per cent increase in demand.

To look at the matter in terms of *proportionate* changes is to bring in the concept of the *price-elasticity of demand*, which is defined as the proportionate change in quantity demanded divided by the proportionate change in price which caused it. For example, if a 2 per cent fall in price causes a 4 per cent increase in quantity demanded, elasticity is 4 per cent ÷ 2 per cent, or 2. If a 1 per cent rise in price causes the quantity demanded to fall by $\frac{1}{2}$ per cent, then elasticity is $\frac{1}{2}$ per cent ÷ 1 per cent, or $\frac{1}{2}$. If a price-change of 3 per cent causes a 3 per cent change in quantity, elasticity is 1, or 'unity' (as the mathematicians would say).¹ The conventional symbol for price-elasticity of demand is ϵ_D .

There is no reason to expect that any given demand curve will have the same ϵ_D along its whole length. Take the case of the downward-sloping straight line in Figure 10.4a (on p. 23 above) which represents Smith's demand. The fact that the slope of the

line is the same at all points means that a price-reduction of (say) 1p always induces Smith to increase his purchases by 2 units of the commodity. But a 1p fall in price is a very small percentage reduction when the price is high to begin with – as it is along the left-hand part of the line near the vertical scale of the diagram; while a 2-unit increase in the quantity demanded will be a *large* percentage increase when the amount demanded is initially very small – as is the case, again, at points on the line close to the vertical scale. This means that ϵ_D will have a high value towards the left-hand end of Smith's curve. Conversely, ϵ_D will be low at points on the curve near the horizontal scale of the diagram; at low prices, a further fall of 1p represents a large percentage cut, while a rise of 2 units in the quantity demanded will be only a trifling percentage increase when the quantity is already large. Somewhere in the middle of the line, there will be a point where the two percentages are exactly equal, so that $\epsilon_D = 1$; to the right of that point ϵ_D becomes smaller and smaller, while to the left it becomes larger and larger, the closer one moves to the extremities.

To a firm which is considering whether to reduce its selling-price, it is only the part of the demand curve where ϵ_D is greater than 1 which is of interest. This is because the firm's total *sales revenue* will not increase in response to a price-cut unless ϵ_D exceeds 1. Suppose, for example, that the firm is currently selling 1,000 units of output per day at a price of 100p, so that its total revenue is £1,000. If ϵ_D is (say) 3, a 1 per cent reduction in price will cause the quantity demanded to rise 3 per cent; in other words, reducing the price to 99p will cause consumers to buy 1,030 units of the product per day, and the firm's sales revenue will now be £1,019.70 (i.e. 99p $\times$ 1,030) instead of £1,000. The firm will now have to decide whether the addition to the total revenue (the *marginal revenue*, as economists call it) is more than enough to cover the extra cost (the *marginal cost*) involved in expanding production by 30 units; if it is, then the firm will find it profitable to reduce its price to 99p. On the other hand, if ϵ_D had been $\frac{1}{2}$, the quantity demanded would only have increased to 1,005 (i.e. by $\frac{1}{2}$ per cent) in response to the 1 per cent price-cut, and total sales revenue would have become 99p $\times$ 1,005, or £994.95; total revenue would have *fallen* from its original level of £1,000, and *marginal revenue* would have been *negative* at -£5.05. Obviously, the firm will do better not to reduce its price in this case – indeed, it would be better business to increase it.¹

In addition to elasticity of demand with respect to price, it may sometimes be important to estimate *income-elasticity* – that is, the proportionate change in demand which results from a small proportionate change in *income*, divided by the proportionate change in income itself; for example, if a 1 per cent increase in buyers' incomes produces a 2 per cent increase in the quantity demanded, income elasticity is 2 per cent $\div$ 1 per cent, i.e. 2. (This, of course, assumes that the commodity's price remains unchanged, and that the increase in demand is due wholly to the rise in incomes.)²

1 When $\epsilon_D = \frac{1}{2}$, a rise of 1 per cent in price (i.e. from 100p to 101p) will cause the quantity demanded to *fall* by $\frac{1}{2}$ per cent (i.e. from 1,000 units a day to 995). Total revenue will thus *rise* from £1,000 to £1,004.95.

2 Still another 'elasticity' concept is 'cross-elasticity' of demand: this is the elasticity with respect to a small proportionate change in the price of some product. For details, see Fleming (1969), p. 82.

7 SUMMARY AND CONCLUSIONS

This unit has attempted to explain how people's economic wants are translated into demands for goods in the market. Because all economic wants are satiable in the short run, the Law of Diminishing Marginal Utility applies – a principle which could be briefly summarised as 'the more you have of anything, the less you want to have still more of it'. Consumers seeking to get maximum satisfaction of wants will allocate their limited spending-power in such a way that the 'marginal' shilling spent on any given commodity gives the same satisfaction as the marginal shilling spent on any other. Because the marginal utilities of some commodities diminish at a faster rate than those of others, consumers may buy fairly small quantities of them even though the *total* satisfaction yielded by them is very great. When individuals' demand curves for a given commodity are summed horizontally, they combine to give 'market' demand curves. The response of demand to price-changes can be defined in terms of the 'price-elasticity' of a market demand curve at any point – an aspect which, as will be seen in the next two units, is also of importance from the point of view of the sellers of commodities.

An obvious limitation of this analysis is that it is much too clear-cut. For example, it assumes that the consumer has complete information about all commodities and their prices. In practice, of course, one of the consumer's abiding problems is to guess the quality of products – for example, to know whether a particular one will really do everything the manufacturer says it will, to estimate how long it will last, and to assess its reliability in use. In our own economy, there are so many different products that it is quite possible for the individual to be unaware that there are acceptable cheaper substitutes for what he is already buying – in which case, he cannot be spending his income to the best advantage.

Another question relates to the development of demand over time. It has been emphasised that demand curves relate to a given period of time, and that they may shift and change from one period to another. The analysis of such changes over time raises a host of new questions – for example, are wants really satiable if the individual has time to adapt his tastes to increasing levels of consumption? Unfortunately, space prevents the investigation of these matters.

Yet another limitation of the preceding analysis is that it appears to relate entirely to the demand for commodities which can be sold through markets. But, as noted in Unit 3, there are some commodities (such as various services rendered by the government) which cannot be provided in this way. In principle, the 'utility' logic must apply to *all* commodities; it should be possible to speak of the marginal utility of (say) street lighting, and to imagine it being compared with the cost of providing it in terms of the other satisfactions lost by diverting scarce resources to this form of 'production'. For reasons both of space and of difficulty, this question will not be pursued any further here – but it deserves to be noted as a limitation on the argument so far.

A question which may also be noted in passing, but which will not be taken any further, is that of the representation of wants

within families and households. Previous references to 'the individual consumer' have implied that the typical buyer of commodities need only consult his own wants before deciding how to spend his income. But many purchases are made by housewives on behalf of their families, by parents on behalf of their children, and so on. Where a buyer has to consider the wants and preferences of a whole group of individuals, trying to allow for the fact that Mrs. Sprat likes fat while her husband does not, the outcome will be more complicated in terms of the degree of satisfaction reached.

Finally, the part played by non-economic wants (as defined in Unit 3, pp. 82-84) has not been considered: yet the desire for the respect and affection of others, or the wish to be 'one up' on one's neighbours, or the longing for mental security, may profoundly affect the pattern of economic wants as such. These questions, however, must be left to the psychologists and sociologists, whose province they are.

To know what everyone wants, and how much people are prepared to pay for things, is not enough to explain why scarce resources are allocated in one way rather than another. As well as the demand side of the market, the supply side must be examined too; and this will be the subject of the next unit - Unit 11.

APPENDIX

I NOTE ON THE MEASUREMENT OF UTILITY (referred to in footnote 3, p. 15)

Economists do not, in fact, have any standard method of measuring utility which could provide the kind of figures shown in Tables 1-3. The argument on pp. 14-17 is not based on evidence from laboratory experiments or systematic observation; it rests partly on an appeal to the personal experience of the reader, and partly on logical deduction from the principle of the satiability of wants. The imaginary figures in the tables are intended to illustrate a chain of *a priori* reasoning, not to present the reader with a set of empirical facts.

However, suppose a laboratory experiment were planned in order to obtain the kind of information given in the tables. It might take the form of presenting a specimen consumer with different quantities of various commodities and asking him to indicate, on a scale of degrees of utility, the amount of satisfaction each of them gives him. But the consumer, no matter how intelligent and co-operative he is, may find it very difficult to assign appropriate numbers to the degrees of satisfaction he experiences; he may, for example, feel that the utility derived from one kind of commodity is of a different nature from that yielded by another, so that it is not possible to compare them or to add them together; and he might object that his enjoyment is not, in any case, the kind of phenomenon which is susceptible of direct measurement in the same way as his height, weight and body temperature.

For these and other reasons, many writers have argued that it is not really meaningful to imagine absolute amounts of utility (or *cardinal* utility) of the kind set out in Tables 1-3. On the other hand, it is still possible to say whether a given set of commodities gives *less* or *more* utility than another, even though the absolute amounts

are unknown; so utility should be treated as an 'ordinal' magnitude instead of a cardinal one.

The laboratory experiment would then proceed by asking the subject to indicate his preferences between alternative sets of commodities. He would presumably always choose larger quantities in preference to smaller ones – e.g. 5 units of commodity A rather than 4 units, or 5A plus 4B rather than 4A plus 3B – as long as he has not yet reached the point of complete satiety. But as between (say) 5A plus 4B on the one hand, and 3A plus 5B on the other, he might prefer the first to the second, or the second to the first, or even find that each gives him the same satisfaction so that he is indifferent which he has; whatever the answer, it will enable the experimenter to rank the two combinations in order of utility. When a large number of such choices has been presented to the subject, the experimenter will have built up what is technically called an 'indifference map', showing his system of preferences but without indicating the degrees of utility yielded by any particular level of consumption.

It turns out, however, that the theoretical results of the 'ordinal utility' approach are much the same as those of the 'cardinal utility' analysis given in the main text of the unit. In place of Diminishing Marginal Utility, it provides a 'Diminishing Marginal Rate of Substitution'. As between any two commodities, the more the consumer already has of A, the less B he will require to compensate him for giving up a single unit of A in exchange. This principle leads to the same conclusions regarding the consumer's behaviour as does the 'Law of Diminishing Marginal Utility'. On the other hand, the exposition of the 'ordinal utility' approach is more complicated and lengthy than that of the more traditional 'cardinal utility' analysis given in the text, and it is for this reason that the latter has been adopted, in spite of possible objections to it on logical and philosophical grounds.

Students who would like to investigate the 'ordinal' or 'indifference curve' approach can do so by reading P. A. SAMUELSON, *Economics: An Introductory Analysis*, Appendix to Chap. 21, or G. J. STIGLER, *The Theory of Price* (3rd edition, 1966), Chap. 4.

MATHEMATICAL NOTES

- M 1 The relationship between the price of a commodity and the quantity demanded can also be stated as an equation. For example, if the quantity demanded (in terms of units per day) is represented by D , and the price (in pence) by P , the equation

$$D = 500 - 10P$$

means that the demand is 500 units a day, *less* ten units for each penny in the price; thus, if the price is 10p (i.e. if $P = 10$), we have

$$D = 500 - (10 \times 10) = 400.$$

Similarly, if the price were 15p, demand would be 350; if $P = 20$ p, $D = 300$; and so on.

If these values were plotted on a graph, it would be found that they form a straight line. Where the price/quantity relationship is a curve, like DD in Figure 10.1 on p. 11, a different type of equation will apply. Unfortunately, equations representing such curves can be quite complicated, so it is sometimes convenient to represent the relationship in a perfectly general way by writing

$$D = D(P)$$

where the ' $D(\dots)$ ' means 'is a function of', and indicates that D depends on P in some systematic way, but without specifying exactly what the relationship is.

On functional relationships in economics, as well as on notation and the mathematical methods used by economists, you may like to consult (whenever you have time) R. G. LIPSEY, *An Introduction to Positive Economics*, Chap. 2 and its Appendix (pp. 17-37).

- M 2 Mathematics students may prefer to think of marginal quantities in terms of derivatives. Thus 'marginal utility' is the first derivative of total utility with respect to quantity, i.e.

$$\text{marginal utility} = \frac{dU}{dQ}$$

where U is utility and Q is the number of units of a commodity. *Diminishing* marginal utility means that d^2U/dQ^2 is negative.

- M 3 A more general definition of the price-elasticity of demand is

$$\frac{\Delta Q_D}{Q_D} \div \frac{\Delta P}{P}$$

where Q_D and P are quantity demanded and market price respectively, and ' Δ ' denotes a small change in the magnitude it precedes. This expression is easily rearranged to give

$$\epsilon_D = \frac{\Delta Q_D}{\Delta P} \cdot \frac{P}{Q_D}$$

(ϵ_D being the symbol for price-elasticity).

BIBLIOGRAPHY

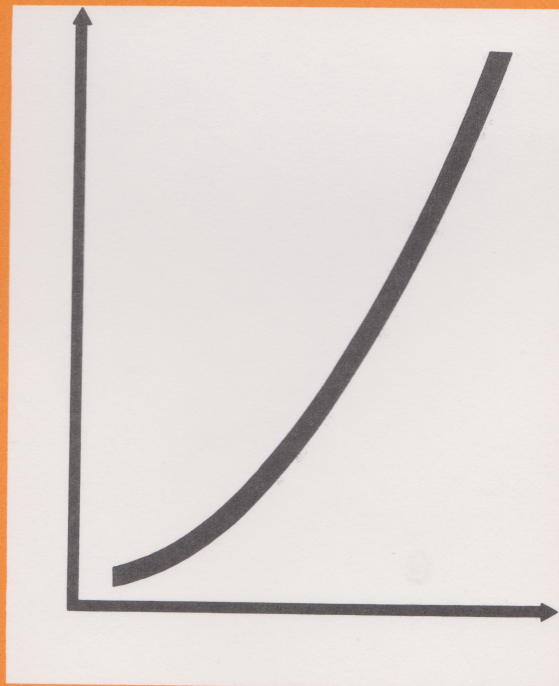
CENTRAL STATISTICAL OFFICE. *Economic Trends* (Monthly). H.M. Stationery Office.

FLEMING, J. M. (1969). *Introduction to Economic Analysis*. London, Allen and Unwin.

SAMUELSON, P. A. (1970). *Economics: an Introductory Analysis*, 8th edition. New York, McGraw-Hill.

Unit 11

Production and Supply



CONTENTS UNIT 11

	PAGE
1 INTRODUCTION: THE SUPPLY CURVE	35
2 THE FACTORS OF PRODUCTION	36
3 INCREASING AND DIMINISHING RETURNS	39
4 THE FIRM AND ITS OBJECTIVES	42
5 OUTPUT, REVENUE AND COSTS	44
6 MARGINAL COST AND PROFIT MAXI- MISATION	47
7 'PERFECT COMPETITION' AND THE SUPPLY CURVE	50
8 THE ELASTICITY OF SUPPLY	52
9 SUPPLY, DEMAND AND THE MARKET	54
APPENDIX	
1. Mathematical Notes	55
BIBLIOGRAPHY	56

PRODUCTION AND SUPPLY

1 INTRODUCTION: THE SUPPLY CURVE

Last week's unit was concerned entirely with the *demand* side of markets, as expressed in demand curves which show how much of any given commodity the buyers will want to buy at each possible price. But this is not enough, by itself, to explain why the price of the commodity should be at one level rather than another, or why any particular quantity should be changing hands between buyers and sellers. For this purpose it is necessary to introduce the idea of *supply* – that is, the willingness of sellers to offer their goods for sale in the market, in quantities which depend on the prices they can get for them. The following example was given in Unit 3:¹

Price (pence per unit)	30	25	20	15	10
Quantity of 'A' supplied (thousand units per day)	200	180	150	100	50

Here, the quantity sellers want to sell each day is an 'increasing function'² of the price of the commodity – i.e. the higher the price, the larger the quantity the sellers will offer on the market, while the lower the price, the less they will want to sell.

Just as Unit 3's demand figures could be depicted as a set of points on a graph (as shown in Fig. 1 of Unit 10), so also can the supply figures be plotted so as to give a *supply curve*. This is shown in Figure 1 below as the curve SS, which passes through all the points indicated by the table. Like the demand curve, the supply curve shows a set of *simultaneous* alternative possibilities, not a succession of actual sales in different time-periods.³ As between one day and another, or one month and the next, the whole curve might shift bodily to right or left – indicating that the sellers' willingness to put their goods on the market has changed, so that in response to a given price they now wish to sell a smaller or larger quantity than before.

Prolonging the curve to the left brings it to the vertical scale somewhere between 5p and 10p, implying that no one is prepared to sell the commodity at all if less than about 7p a unit can be obtained for it – because (for example) even the cheapest producer would be so far from covering his costs at this price that he would find the commodity no longer worth making. Prolonging the curve to the right causes it to slope more and more steeply until it is

1 See Unit 3, page 73.

2 The idea of a 'function' was briefly explained in the Mathematical Note M 1, in Unit 10. To say that y is an 'increasing' function of x means that y increases whenever x does; if it were a 'decreasing function' it would get *smaller* as x got larger.

3 See Unit 10, pages 12–13. It will be seen later that there is an additional time-factor on the supply side, which has to do with the speed with which producers can adapt to different levels of production.

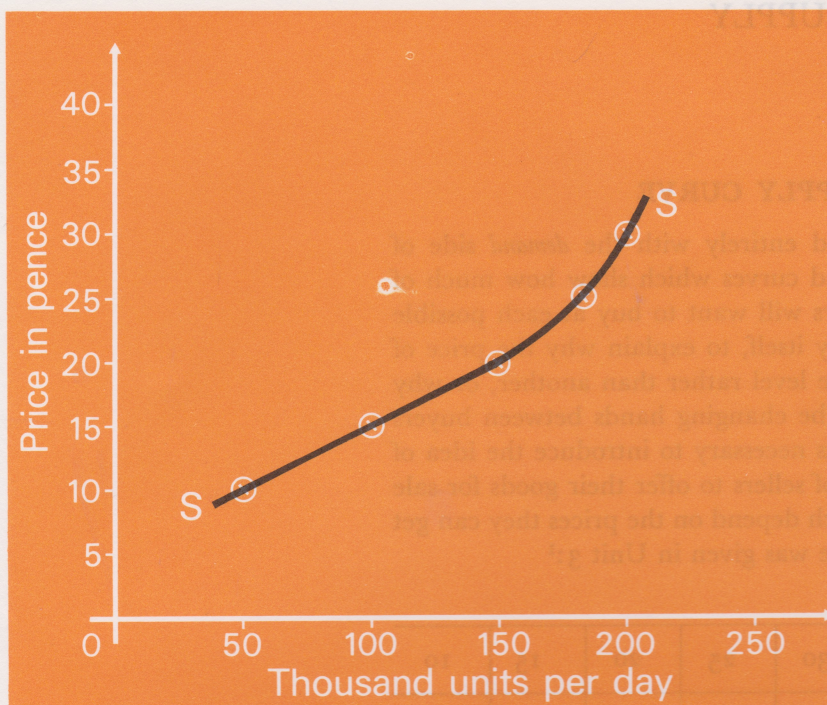


Figure 11.1

eventually vertical; this means that however high the price may rise, it will not induce supply to exceed some maximum quantity, because producers will have reached the limit of their capacity for the time being.¹

To explain what determines the shape and position of any supply curve, it is necessary to study the behaviour of the firms engaged in producing and selling the commodity. Strictly, 'supply' means the willingness to *sell*, rather than to *produce*; it is quite possible for a firm to sell without producing, if it has a previously-accumulated stock of goods on which it can draw; similarly, it can produce without selling, by adding its current output to stock instead of immediately putting it on the market. On the other hand, it is obvious that sales could not continue very long without exhausting the stock of most commodities, so that supply would then be impossible without further production; while the impossibility of accumulating stocks indefinitely means that current production must sooner or later be offered for sale. Though production and supply are not the same thing, they are so intimately connected that it is not possible to explain one without the other.

2 THE FACTORS OF PRODUCTION

Production is the transformation of *inputs* of labour, materials, and the service of capital equipment, into *outputs* of goods and services which satisfy wants in some way. Some outputs satisfy wants directly, by being eaten or worn, or otherwise used up by consumers; others do so indirectly by serving as inputs in subsequent production processes – as, for example when steel (an 'output' from the point of view of the iron and steel industry) is used as an input in the motor-car industry. When the economy is taken as a whole, these inter-industry exchanges cancel one another out; they are like the

passing-on of semi-finished products from one branch to another of a single large enterprise.¹ The inputs which do *not* cancel out are the services of the *factors of production*, which are conventionally grouped into three main categories:

- (a) human labour-power, both mental and manual;
- (b) 'land' – here used as an inclusive term for all natural resources, such as mineral deposits, shoals of fish, and the hydro-electric potential of rivers as well as the soil itself;
- (c) 'capital', i.e. the means of production which have themselves been produced; the greater part is 'fixed' or 'durable' capital, such as plant and machinery, ships, vehicles and tools of all kinds, while the remainder consists of stocks of materials.² (The process of *adding* to the existing quantity of capital, by building new plant and equipment or accumulating stocks, is called by economists *capital formation* or 'investment'.)³

This classification emphasises the distinction between the means of production which exist ready-made, so to speak, within mankind itself and in its environment (i.e. labour-power and land), and those which come into existence as a result of human effort (i.e. capital). For capital to be built up and maintained, it is necessary that some part of the economy's current output should take the form of machines, buildings, etc. instead of goods and services for immediate consumption; the community must be prepared to abstain from current consumption to the necessary extent – or (to put it another way) to postpone consumption by waiting until the newly-constructed capital equipment eventually increases the output of consumables. Thus, it can be said that physical capital is merely the embodiment of the *abstinence* or *waiting* which is the ultimate factor designated by 'capital'.

To the three groups of factors just mentioned, many economists add a fourth, namely *enterprise* (or *entrepreneurship*).⁴ This is the contribution made by *firms* to production by organising the processes through which the services of the other factors are co-ordinated to yield outputs. Firms (defined in the very broad sense adopted in Unit 3, which ranges from small one-man businesses to giant

1. Except for materials which are imported from other countries. If the national economy were regarded as a single giant firm, its imports would be its inputs of materials from similar 'firms' overseas, and would not cancel out. For simplicity, it will be assumed that the economy is self-contained, so that this point can be ignored for the time being. It will come up again in Unit 14.

2. Students who consult American textbooks, such as Samuelson (1970), will find that the word 'inventory' or 'inventories' is used for 'stocks'.

3. Both 'capital' and 'investment' have other meanings in everyday speech. 'Capital' can be used to mean a holding of shares or bonds or some other kind of 'paper asset', while 'investment' often has the sense of buying shares or bonds. Here, the words refer solely to *physical* assets like machines and buildings. In later units, too, 'capital' and 'investment' will normally be used in the 'physical' sense unless otherwise indicated.

4. In the early nineteenth century, English economists used the word 'undertaker' to refer to the organiser of a business, but this was abandoned later in the century in favour of the French word '*entrepreneur*'—which has the same economic meaning but has no suggestion of funerals to the English ear. It would have been simpler to coin a new English word 'enterpriser'—but for some reason this was not done, and '*entrepreneur*' became one of the technical terms of English economics.

companies¹) initiate new projects, furnish the managerial 'know-how' which gets them running, and bear some of the risks involved; and it can be argued that this 'entrepreneurial' function is sufficiently distinct, both from labour and from the 'abstinence' which provides capital, to justify separating it out as a fourth factor of production. The same principle might even be applied to *technology*, which would then count as a fifth factor. It is true that it is usually inseparable from other factors, being 'embodied' in the design of machinery and the skills of workers; all the same, firms often plan and finance research programmes to devise new technologies, just as they invest in the construction of new physical capital; and it is possible to buy and sell technology, as (for example) when one firm sells another the right to produce a patented commodity under licence, or to use a special process which it has discovered.

A rather different approach, which avoids these problems of classification, is to regard the process of production as a *function* in which the output of any commodity (Q) depends on the inputs of factor services (F_1, F_2 , etc.) in accordance with a functional relationship (ϕ), as follows:

$$Q = \phi (F_1, F_2, F_3 \dots \dots)$$

In this *production function*,² F_1 represents man-hours of a particular type of labour, F_2 'machine-hours' provided by a particular type of capital equipment, and F_3 'acre-hours' of a certain grade of land; while the function sign $\phi(\dots \dots)$ represents the method by which they are combined together to give the output Q – that is, it stands for the organisation and technology discussed in the previous paragraph. Given ϕ , any increase in F_1, F_2 or F_3 , or in all of them together, will cause Q to increase; while an improvement in technology, which causes ϕ to change, will increase Q even if F_1, F_2 , etc. have remained exactly the same as before.

Here, the inputs of factor-services F_1, F_2 , etc. must be defined in such a way that each of them consist of *homogeneous* units – i.e. units which are perfect substitutes for one another. Thus, F_1 could be said to represent man-hours of 'labour' only if all labour were exactly alike. In practice, of course, the skills and abilities of (say) coal-miners, shop-assistants and aircraft designers are so different that an hour of a miner's labour cannot be meaningfully added to an hour of a shop-assistant's; they must be treated as the services of separate factors of production. The same thing applies to the different varieties of capital: lathes, electronic computers and printing-presses are obviously so very dissimilar, both in themselves and in the services they provide, that they must be counted as different factors. So the total number of factors in the economy as a whole is very large when they are precisely defined, and the threefold

1 See Unit 3, page 71.

2 This is a well-established technical term in economics. A great deal has been written on the concept of the production function, but it is not possible to go more deeply into it here. On functions in general, look back to footnote 2 on page 35 above, or refer to Mathematical Note **M** 1 at the end of Unit 10. Remember that the term ϕ merely indicates that there is a systematic relationship between inputs and outputs; it does not specify the relationship itself. The whole equation is to be treated as no more than a piece of mathematical shorthand, saying 'the quantity of output depends in some systematic way on the amounts of various factors whose services are used as inputs into the production process'.

'labour-land-capital' division is seen to be only a very rough grouping. Some factors, indeed, will be hard to assign to any one of the three categories, because they represent a mixture of more than one: for example, irrigated land is more than just a 'free gift of nature', since its fertility is largely the result of work done in the past (i.e. investment)¹ on the construction of channels and sluices; it is therefore a blend of both land and capital.

Labour, too, includes an element of capital: where skills and abilities are the result of previous training and education, they can be said to be 'produced means of production', and the use of scarce resources in the training process can be regarded as 'investment in human capital'. Thus, the crude 'labour-land-capital' classification may even be misleading if it is taken to mean that all factors must fall neatly into one or other of the three categories. It is sometimes convenient to assume, as a way of simplifying theoretical arguments, that there is only one kind of labour, only one type of machine, and only one kind of land; but it must always be remembered that this is a gross over-simplification of reality.

In the 'production function' given earlier, the F's stood for the *services* of the factors, not the factors themselves. It is important to keep this distinction in mind. The factors are the quantities of labour-power, machinery, etc. which exist at any given moment of time; they are *stocks* of assets which are there whether they are put into use or not. Their services, on the other hand, are *flows* which occur over time, and can only be measured with reference to some particular time-period, for example as 'man-hours', 'man-days', 'machine-hours', 'acre-years' and so on.²

TEST 1

3 INCREASING AND DIMINISHING RETURNS

In most production processes, the different factor-services are usually *complementary* – that is, they must be used in conjunction with one another if they are to produce anything at all.³ Machines can produce nothing without labour to operate them;⁴ workers can make nothing with their bare hands;⁵ and without an input of materials provided by natural resources (i.e. 'land' in the broad sense) there can be no output. In some cases, the technique of production makes it necessary to use factor-inputs in *fixed proportions* to one another. If a certain kind of machine needs exactly five workers to operate it, the input of man-hours must always be exactly five times the number of machine-hours; to employ, say, twelve men with two machines will mean that two of the men are permanently idle, while to instal two

1 In the 'physical' sense given on page 37 above.

2 The question of stocks and flows came up in Unit 10, page 12, footnote 3. You may find it useful to refer back.

3 In Unit 10 (p. 21) goods were said to be 'complementary' if they were closely associated in *consumption*. The word is being used in exactly the same sense here, except that it is *factors* which are closely associated in *production*.

4 There are some obvious exceptions to this: for example, 'automotive' plant is capable of operating unattended.

5 Workers can, however, render certain *services* without equipment or materials; again, this statement must be regarded as subject to exceptions.

machines when only five workers are available will mean that the second machine cannot be used and will produce nothing. But complementarity is not always so rigid as this. In agriculture, for example, it is possible to grow crops by labour-intensive methods which concentrate a great deal of labour on each acre, or alternatively by land-intensive methods which distribute the available labour over a great deal of land. Here, the fact that the services of the factors can be combined in *variable proportions* means that it is possible, to some extent, to substitute one for another in the production process, as shown in the following example.

Suppose a farmer is growing wheat on 100 acres and employing 10 men for the purpose,¹ so that the existing proportion between factor-inputs is 10 acre-years to 1 man-year. He now wishes to increase output by 10 per cent. He could do it by taking on one more man and renting 10 additional acres; this would preserve the original land/labour ratio of 10 to 1. But now let us suppose that factor-proportions are variable, and that an alternative way to increase output by 10 per cent is to rent no additional land at all, but to employ *three* more men to cultivate his existing acres more intensively. He can then achieve the same objective *either* (a) by taking on three more men and no extra land, *or* (b) by employing one more man and renting ten additional acres. It follows that two of the three men he would be employing under (a) are equivalent, in terms of their effect on production, to the ten acres he would be renting under (b). In other words, the services of labour can be substituted for those of land at the rate of one man-year for five acre-years.

However, land and labour are not *perfect* substitutes for one another. If they were, the farmer could give up *all* his original 100 acres, take on twenty additional workers in place of them, and produce his original output with 30 men and no land at all! Alternatively, he could increase the output of his 100 acres to any desired level merely by taking on proportionately more labour, so that instead of needing 20 men and 200 acres to produce twice his original crop, he could do it by employing 40 men to cultivate the original 100 acres; he could produce ten times the original amount with 280 men, and a hundred times the amount with 2,980 men, all working on the 100 acres he started with – which is manifestly absurd.² The fact is that, as the proportion of labour to land increases, labour's efficiency as a substitute for land will gradually diminish, and more and more of it will be needed to raise production by a given amount. Eventually, a level of output will be reached for which 100 acre-years is an indispensable minimum input, which no amount of labour, however large, can possibly replace. The substitutability of labour for land will have fallen to zero, and no further increase in output will be possible unless the quantity of land is increased.

This exemplifies the *Law of Diminishing Returns*, which says that if all but one of the factor-inputs in a production process are kept fixed, successive equal additions to the remaining factor-input will

1 These figures are used only for the purpose of providing a simple numerical example: they do not claim to be in any way realistic. Students who happen to be farmers are asked to exercise extreme patience here.

2 A standard textbook argument is as follows. If, when more and more labour is applied to a given area of land, the output grew in exactly the same proportion, it would be logically possible to grow the entire world's food supply in one window-box.

eventually cause output to increase by smaller and smaller amounts. The word 'eventually' allows for the possibility that there may be *Increasing Returns* in certain cases; that is, where the amount of the variable factor-input is initially very small relatively to the fixed inputs, additions to the variable input may for a time cause output to grow by successively *larger* amounts. But this cannot continue indefinitely, and a point must inevitably be reached at which diminishing returns set in. In Figure 11.2 below, the curve OW illustrates these conditions. The horizontal scale OF measures the input per time-period of the variable factor; the vertical scale OQ measures total output per time-period. Between O and A, the slope of the curve is steadily increasing, showing that additional units of the variable factor cause output to rise by successively larger amounts. From A to B, however, returns are diminishing; the proportion of

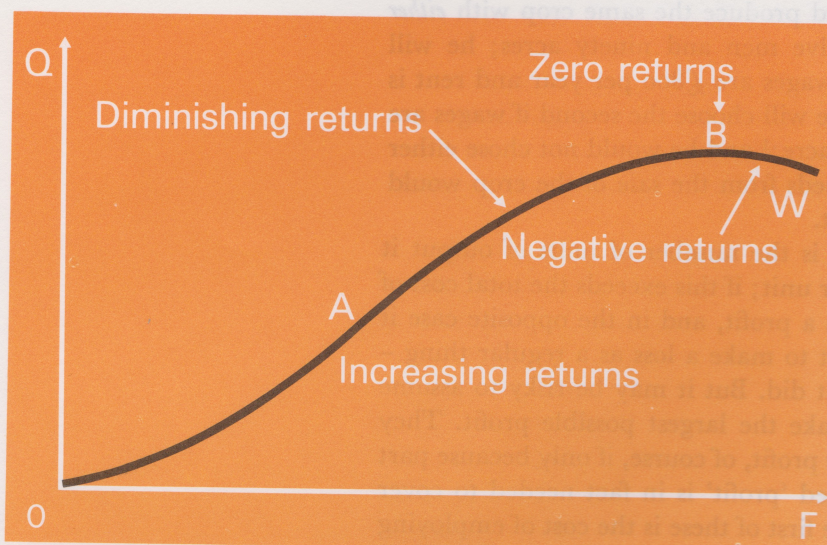


Figure 11.2

the variable factor-input to the fixed factors has now become so large that its substitutability is decreasing. At B, maximum output is reached, and an additional unit of the variable factor will yield no extra output at all; at this point, a condition of 'zero returns' exists. Finally, between B and W there are 'negative returns': the amount of the variable factor has now become so great that further units actually *reduce* total output – as, for example, when the number of workers on a given acreage of land has become so large that they are getting in one another's way and trampling down one another's crops.

It should be emphasised that the conditions shown in Figure 11.2 assume a given technique of production. If a new and better technology were introduced, under which the whole process became more productive, the effect would be to shift the whole curve vertically upwards. Its starting-point would still be at O (the 'origin') but all other points would move upwards in proportion to their distance from the horizontal. The same effect would occur if the fixed inputs themselves were to be increased, so that each quantity of the variable input now had more of the others to work with. Finally, it should be noted that Figure 11.2 describes the response of output to changes in only *one* factor. If *all* inputs were increased together in the same proportion, the response of total output could be described as *Returns to Scale*. If the simultaneous doubling of all inputs caused output also to double, this would be *Constant Returns to Scale*; if output more than doubled, there would be *Increasing Returns*

to Scale; while *Diminishing Returns to Scale* would mean that output rose *less* than proportionately as the factor-inputs were increased.

TEST 2

THE FIRM AND ITS OBJECTIVES

From the point of view of the *firms* which organise production by bringing the various factor-inputs together, all these physical relationships are of great importance. But they are far from being the whole story, because the managers of firms must also pay careful attention to the *costs* of employing the factors, and compare these costs with the *revenue* to be obtained from selling the output. For example, suppose a farmer could produce the same crop with *either* ten men and 100 acres *or* twelve men and ninety acres; he will choose the first combination if wages are £800 per year and rent is £120 per acre per year, while he will choose the second if wages are still £800 but land is £200 an acre;¹ and he would not choose either of them if he thought the proceeds from the sale of the crop would not cover the costs of growing it.

The revenue a firm receives is the number of units of output it sells, multiplied by the price per unit; if this exceeds the total cost of producing the output, it makes a profit, and in the opposite case it makes a loss. No firm will want to make a loss as a regular thing – indeed, it could not survive if it did. But it may be risky to assume that firms invariably try to make the largest possible profit. They will certainly wish to make *some* profit, of course, if only because part of what is conventionally termed ‘profit’ is in fact needed to cover certain costs of production. The first of these is the cost of employing *capital*. If a firm were to rent its factory buildings from a landlord, and hire all its machinery from a company which specialised in the renting-out of equipment, the expense incurred under these headings would appear in its accounts quite clearly and unambiguously as a cost of production. The fact that most firms are outright owners of the equipment they use does not mean that it can be regarded as costless. The owner of a small business, for example, might reasonably reckon the ‘cost of employing capital’ as the income he would have received if, instead of using his money to buy plant and equipment, he had lent it out at interest or made some other gainful use of it. A large company, whose equipment was bought with borrowed money, could regard the annual interest-cost as the equivalent of what it would have had to pay to hire its plant and machinery direct. It will be useful to refer to this ‘cost of capital’ as *capital rental*, even though, for most enterprises, it does not represent an amount of money actually paid out by the firm to someone else.

The second of the cost elements in ‘profit’ is what may be called ‘minimum profit’ (or, as some writers call it,² ‘normal profit’). Suppose an ‘entrepreneur’ has organised a production process, employing all the necessary labour, renting the buildings and equipment, obtaining raw materials on credit, and so on; no capital of his own is employed, and even the day-to-day administration of

1 Once again, it is emphasised that these figures are not claimed to be at all realistic: they are merely intended to provide a simple numerical example.

2 See, for example, Fleming (1969), p. 116; Samuelson (1970), p. 457.

the enterprise is in the hands of salaried managers. Here, it might seem, all costs of production are being allowed for before profit is calculated; yet it would still not allow for the fact that, unless some minimum amount is forthcoming to the entrepreneur himself, he will not think it worth his while to organise the enterprise in the first place and keep it running thereafter. Without the prospect of at least this minimum of profit, production would not be undertaken – in just the same way as a failure to pay wages would cause it to stop for lack of labour, or the non-payment of bills for materials would halt output for lack of supplies.

Any firm, therefore, will try to make at least this minimum amount of profit, in addition to the capital rental already mentioned. On the other hand, the assumption that firms try to *maximise* profit implies that they will not be content with the minimum if they can make more, and that they will decide on their levels of output, the prices of their products, and so on, with no other consideration in mind but that of making the highest possible profit. In fact, however, firms often do allow other objectives to take precedence. The uncertainties and risks of business life may induce them to set stability above profitability, for example by refraining from raising prices to take advantage of a rise in demand which may turn out to be short-lived. An entrepreneur who is interested in engineering for its own sake may neglect profitable markets in which simple products can be sold, in favour of less promising markets for technically more sophisticated ones. Some firms believe in being ‘good employers’ on general social grounds, and provide benefits and amenities whose costs are not strictly necessary as labour incentives. Often, too, firms do things which are unprofitable in the short run in the hope that they will pay off in the long – for example, selling current output at a loss in order to stimulate demand in the future; this raises the question of how ‘profit-maximising behaviour’ is to be defined and identified in the first place!

In practice, then, the assumption that firms always strive to maximise current profits is more complicated and questionable than might at first appear. But as a broad simplification, it is not at all unreasonable; most business men would agree that profit-making is their main objective, even if it is not the only one. It can, indeed, be argued that it is in the public interest that it should be so. If prices are assumed to reflect the degree to which commodities satisfy people’s wants, then the fact that profit-making entails the transformation of relatively lower-priced inputs into higher-priced outputs means that total satisfaction is being increased thereby; the higher the level of profit, therefore, the greater the excess of output value over input value, and the greater the contribution made by firms towards bringing about a desirable allocation of resources.¹ In any case, it will be very convenient to make the

1 In the contrary case, where a firm makes a loss because it is unable to sell its output at a price which covers production costs, it can be said to have converted valuable resources into commodities which are *less* valuable – a proceeding which is undesirable from society’s point of view. This point may, perhaps, strike you as a highly ‘capitalistic’ argument. If so, refer back to Unit 3, pages 81–82. Under ‘market socialism’, profits would be the aim of the publicly-owned enterprises which were responsible for production. In a ‘command economy’, too, it would also be a social objective to transform commodities from *less* want-satisfying forms into *more* want-satisfying ones, even though market prices were not available as indicators of the satisfaction given by the different goods and services.

'profit-maximisation' assumption in the next stage of the argument, which concerns the basic principles on which a firm should determine its level of output in order to make as much profit as possible.

5 OUTPUT, REVENUE AND COSTS

The greater the output a firm produces during any period of time, the greater will be its costs of production; the greater, too, will be the total revenue (i.e. quantity sold $\times$ price) unless the elasticity of demand is less than unity.¹ For the time being, it will be assumed that the firm can sell any amount it likes at an unchanging price, so that total revenue depends only on the quantity sold; thus, if the price of the firm's product is £1, an output of 200 units a day will yield a total revenue of £200, output of 300 units will yield £300, and so on. The implications of this will be explored later on; at present, the assumption is made purely as a means of simplifying the problem.

If production cost were also a constant amount per unit – say, 75p – the firm's best policy would obviously be to produce and sell as much as it possibly could; total profit would be the number of units of output multiplied by 25p. But there are reasons why, in the short run at least, cost per unit will *not* be constant, but will itself vary according to the level of output. Thus, if cost per unit (i.e. *average cost*) is £1.25 at an output of 100 units, 75p at 200 units, and £1.50 at 300 units, the firm will make a profit only if its output is in the neighbourhood of 200 units. To see why such variations should occur, it is necessary to consider production costs in more detail.

An important distinction is between *fixed* and *variable* costs,² the former being those which do not vary with the level of output, while the latter are those which *do* so vary. *Fixed costs* include such items as the 'capital rental' of the firm's plant and equipment; whether machinery is kept working at full capacity, or held idle and producing nothing, its cost per period will be much the same. The wages of the office staff and of night-watchmen, the upkeep of buildings, and local rates provide further examples; the 'minimum profit' discussed earlier must also be included as a fixed cost – production will not be undertaken and maintained without it, but its amount does not depend on the level of output. *Variable costs* include the cost of materials used, since this will clearly be proportionate to the volume of output being produced; fuel, too, is likely to be required in accordance with the level of output. The wages of 'shop-floor' labour are usually counted as a variable cost, on the ground that the size of the labour force, and its hours of work, can be adjusted fairly quickly to the level of output to be produced; however, this is subject to some qualifications which will be noted presently.

Since fixed costs are a constant amount which remains the same whatever the level of output, *average fixed cost* (i.e. per unit of output) will be smaller, the larger the number of units produced. For example, if they are £100 a day, average fixed cost will be £50 when two

1. On this point, you may want to refer back to the explanation of 'elasticity' in Unit 10, pages 24–5.

2. 'Fixed' costs are often referred to as 'overhead' or 'supplementary' costs; 'variable' costs are also called 'running costs' or 'prime costs'.

units are produced, £20 when output is five units, £5 when it is 20 units, and so on. The relationship between output and average fixed cost is shown by the curve AFC in Figure 11.3 below. Its shape indicates that this element of cost becomes smaller and smaller per unit as total output increases.

The curve marked AVC shows the amount of *variable* cost per unit at each level of output. Its shape suggests that, as more materials, labour, etc., are used in conjunction with the fixed amount of plant and equipment the firm possesses (and whose costs are represented by AFC), 'increasing returns' operate up to an output of about 10 units a day; then, as output becomes larger, 'diminishing returns' set in – larger and larger amounts of the variable factors (labour and materials) are needed per unit of output, so that average cost necessarily increases. On the extreme right, the curve climbs very steeply; if extended, it would become vertical, indicating that output cannot be made to exceed a certain figure (about 28 units a day) no matter how much is spent on increasing the variable inputs of the production process.

The total production cost per unit is the sum of the fixed and variable costs, as shown by the curve $A(F+V)C$ in Figure 11.3: the height of any point on this curve is the sum of the heights of AVC and AFC vertically below it. Its U-shape is due to the fact that at low levels of output (say, below 10 units a day) fixed cost per unit becomes very large, while at high levels of output (say above 25 units) variable cost per unit rises sharply as diminishing returns become marked. If the firm were to extend its plant and thereby increase the scale of its operations, the AFC curve would be pushed outwards from the 'origin' of the diagram, and the AVC curve would be shifted to the right, with its lowest point corresponding to a higher output than before; thus, the combined $A(F+V)C$ curve would be pushed rightwards. An increase in variable cost alone (e.g. because of an increase in the price of materials) would merely push the AVC and $A(F+V)C$ curves upwards.

It should be noted that the division of costs into 'fixed' and 'variable' is not a rigid one; it depends on the time-period which is assumed. Over (say) five years, a firm could replace a good deal of its machinery; it could let it run down, or add to it, according to the quantity of output it expected to produce in future; thus,

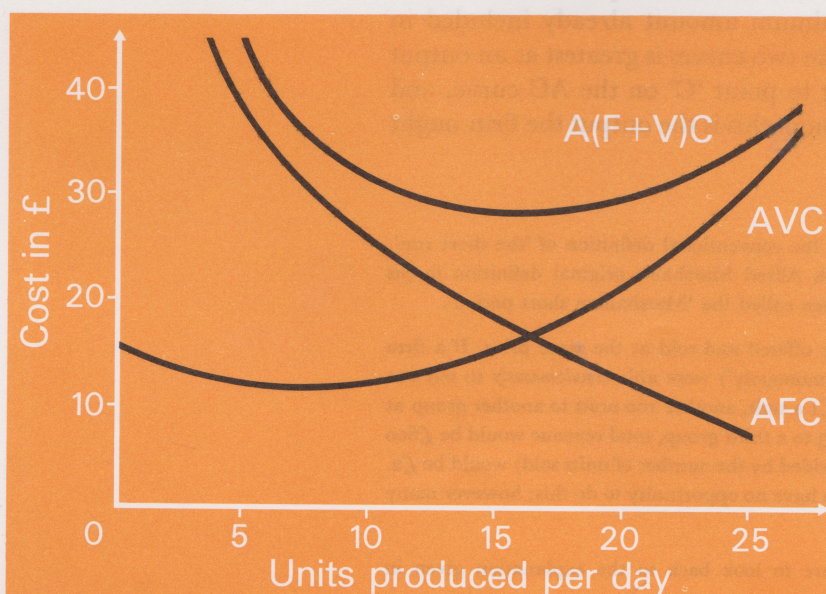


Figure 11.3

'capital rental' would count as variable cost rather than fixed. At the other extreme, a one-day time-period would mean that relatively few costs were variable; in many enterprises, for example, it would be impossible to cut down the labour force for one day even if production were to be considerably reduced. So the longer the time-perspective, the smaller the proportion of costs which are fixed, and the greater the proportion which are variable; the AFC curve of Figure 11.3 will lie closer to the vertical and horizontal scales, and the AVC curve will be shallower as well as higher; the $A(F + V)C$ curve, too, will have a shallower curvature. In the *very* long run, the AFC curve will vanish altogether, and the AVC curve will approximate to a horizontal line. In what follows, however, it will be assumed that the time-period is too short to permit the firm's capital equipment to be significantly increased or decreased;¹ 'capital rental' is thus a fixed cost, and the $A(F + V)C$ curve is shaped more or less as shown in Figure 11.3.

The possibilities open to the firm can now be seen in Figure 11.4, which compares cost with revenue per unit of output. The 'AC' (average cost) curve is the same as Figure 11.3's $A(F + V)C$; the 'F + V' has been omitted because the separate AFC and AVC curves do not appear. The horizontal line marked ' $P = AR = MR$ ' shows the amount the firm can expect to receive for each unit of output it sells. It has already been assumed that the firm can sell any amount it likes at a given price; here the price ('P') is £25, and the fact that the line is horizontal means that P is the same for all quantities shown on the horizontal scale. Price is the same thing as 'average revenue', AR,² and in this case it is also identical with *marginal* revenue, MR – i.e. the addition to total revenue which occurs when the firm sells an extra unit of its product.³ Wherever the AC curve lies *above* the ' $P = MR = AR$ ' line, the firm will be operating at a loss, since the price per unit is less than the production cost per unit; thus, it will not wish to produce less than 7.5 units per day, nor more than 23.5 units. At points 'A' and 'B', where the lines cross, the price is just enough to cover average cost; since the latter includes both 'capital rental' and 'minimum profit', the firm can afford, if necessary, to produce either of the levels of output indicated by 'A' and 'B'. But it would obviously do better to choose an output *between* the two points, since the AC curve then falls well below the ' $P = MR = AR$ ' line, indicating that a substantial profit can be made per unit in addition to the minimum amount already included in average cost. The gap between the two curves is greatest at an output of 15 units a day, corresponding to point 'C' on the AC curve, and it might seem at first sight as though this is the output the firm ought

1 In the literature of economics, this is the conventional definition of 'the short run'. Because it corresponds more or less with Alfred Marshall's original definition in his *Principles of Economics* (pp. 374–7) it is often called the 'Marshallian short period'.

2 Provided, of course, that all units are offered and sold at the same price. If a firm (of the type known as a 'discriminating monopoly') were able *simultaneously* to sell 100 units a day to one group of customers at £1 a unit, another 100 units to another group at £2 a unit, and still another 100 units at £3 to a third group, total revenue would be £600 and average revenue (i.e. total revenue divided by the number of units sold) would be £2. But the firm discussed above is assumed to have no opportunity to do this; however many units it sells, each is priced at £25.

3 On 'marginal revenue', you may care to look back to the explanation given in Unit 10, page 25.

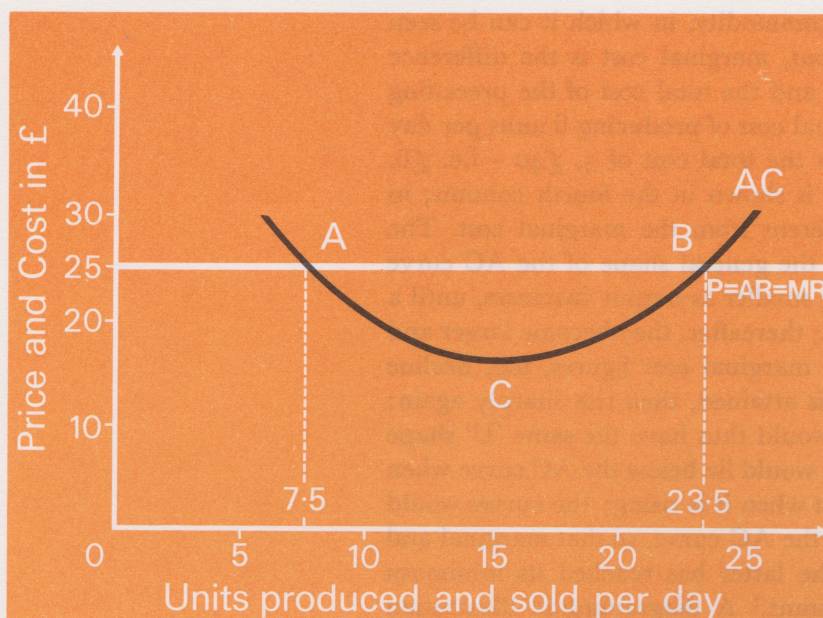


Figure 11.4

to choose if it wants to maximise profit. But it is one thing to maximise profit *per unit*, another to maximise *total* profit; it is better to sell 100 units at a profit of 4p a unit, than to get a profit of 5p a unit on sales of 75 units a day. In this case, the firm will find that its total profit will increase if it produces and sells *more* than the 15 units a day indicated by 'C', and accepts a smaller amount of average profit. But how can it determine the exact level of output which will maximise total profit? To answer this question, it is necessary to introduce the concept of *marginal cost*.

6 MARGINAL COST AND PROFIT MAXIMISATION

The concept of 'marginal' quantities was first introduced in Unit 10, when it was applied to utility. 'Marginal utility' was defined as the addition to total want-satisfaction brought about by giving the consumer an additional unit of a commodity.¹ In just the same way, *marginal cost* is defined as the addition to total cost which is brought about when the production of a given commodity is increased by one unit.² The table below gives a numerical example of the pro-

Number of Units per day (a)	Total Cost (£) (b)	Marginal Cost (£) (c)	Average Cost (£) (b) ÷ (a)
1	18	—	18
2	28	10	14
3	33	5	11
4	36	3	9
5	40	4	8
6	48	8	8
7	63	15	9
8	80	17	10
9	108	28	12

¹ See Unit 10, page 15.

² M 2.

duction costs of a hypothetical commodity, in which it can be seen that, at any given level of output, marginal cost is the difference between that output's total cost and the total cost of the preceding level of output. Thus, the marginal cost of producing 6 units per day is the total cost of 6, £48, minus the total cost of 5, £40 – i.e. £8. The *average* cost of each output is shown in the fourth column; in every case except one it is different from the marginal cost. The average cost figures conform to the general shape of the AC curve of Figure 11.4, at first becoming smaller as output increases, until a minimum value of £8 is reached; thereafter, they become larger and larger as output increases. The marginal cost figures, too, decline at first until a minimum of £3 is attained, then rise sharply again; a curve showing marginal cost would thus have the same 'U' shape as Figure 11.4's AC curve, but it would lie below the AC curve when the latter is falling, and above it when it is rising; the curves would intersect at the lowest point on the AC curve, so that marginal and average cost are equal when the latter has reached its minimum value and is momentarily constant.¹ A diagrammatic illustration will be provided presently; if you are impatient to see what the curves look like, turn to Figure 11.5 below.

The significance of marginal cost can easily be appreciated if the firm is assumed to be able to sell its product at £16 a unit. The output at which profit *per unit* is highest is six units; with cost per unit at £8, the firm will make a profit of £16 minus £8 on each unit it produces and sells, and *total* profit will be £48. But it could do better still by selling seven units; the seventh unit will increase total

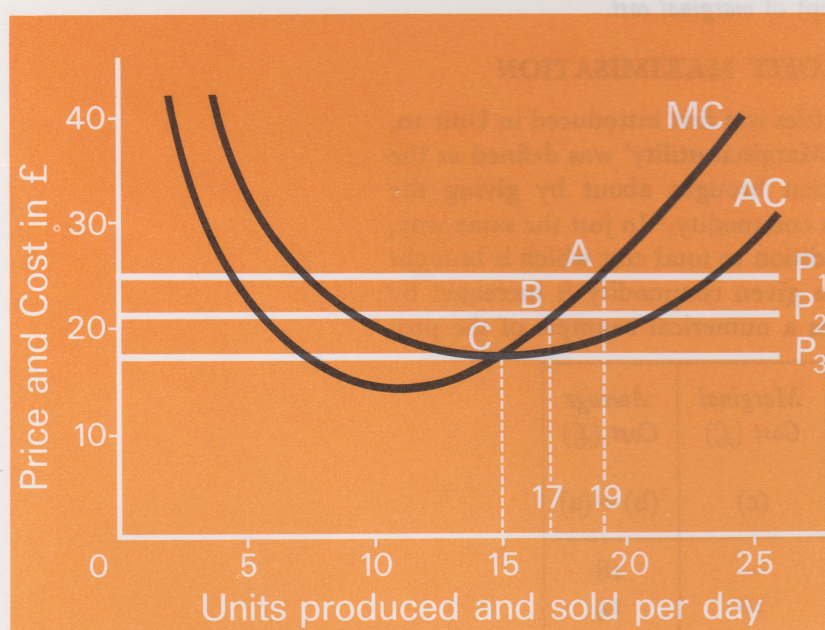


Figure 11.5

1 The relation between marginal and average cost has often been likened to a sequence of scores in a cricket match. Suppose five batsmen have between them scored 150 runs, so that the average is 30. If a sixth man now comes in and scores 30, his 'marginal' score (i.e. his addition to the total runs scored) will be exactly equal to the average. To *raise* the average to 32, he would have to score 42 – i.e. he must not only score 32 himself but an additional two runs for each of his predecessors; so his 'marginal' score – 42 – must be higher than the increasing average of 32. On the other hand, a 'marginal' score of 18 will reduce the average to 28; when the average goes *down*, the marginal score must be *below* it. (It should, perhaps, be added that the fifth batsman, who will still be in, is assumed not to score any runs himself during this episode.)

revenue by £16, but (as shown by the marginal cost figure in the third column of the table) its production will add only £15 to total cost; consequently, there will be an addition of £1 to total profit, which will now be £49. On the other hand, to produce and sell eight units instead of seven would increase total cost by £17, i.e. by £1 more than the 'marginal revenue' of £16 which would accrue from the sale of the eighth unit; so the firm should confine itself to producing seven units per day.

The general rule to be deduced from these calculations is that the firm should increase its output as long as the sale of an additional unit will add more to total revenue than it does to total cost – that is, as long as marginal cost falls short of marginal revenue; it should reduce output if marginal cost exceeds marginal revenue; and it should neither increase nor reduce output when marginal revenue is equal to marginal cost, since this will be its most profitable position.¹ In the case of the firm which has been under consideration so far, marginal revenue is the same as the price of the product, since this has been assumed to remain unchanged whatever the level of output the firm puts on the market. So the rule for maximising profit – i.e. *produce the output at which marginal revenue equals marginal cost* – can in this case be changed to *produce the output at which marginal cost equals the price of the product*. The implications for firms whose marginal revenue is normally less than price will be examined in Unit 12.

Meanwhile, the effect of changes in the market price on a firm's output can now be seen with the aid of Figure 11.5. This is the same as Figure 11.4, except that it now includes a curve marked 'MC' which shows the amount of marginal cost at the various levels of output. The original 'P = AR = MR' line at £25 is now marked P₁, and two similar lines marked P₂ and P₃ have been added. For the moment, ignore P₂ and P₃, and suppose that P₁ is the ruling market price as it was in Figure 11.4. Now that the marginal cost curve 'MC' has been inserted, it can be seen that the output which will give the firm maximum profit is 19 units per day, because this is the output whose marginal cost is equal to the price of the product – as shown by the fact that the MC curve intersects the line P₁ at point 'A' vertically above the '19' mark on the horizontal scale.

But now suppose the market price of the commodity falls to the level indicated by the line P₂ – i.e. £22. The fall in price is not due to any action taken by the firm itself; it is simply a question of conditions having altered in the market as a whole. The firm can still sell as much as it likes at the ruling price, except that the latter is now £22 instead of £25 as before. At the original level of output, 19 units, marginal cost will now be in excess of marginal revenue, and the firm will therefore reduce output to the point where the MC curve intersects P₂, i.e. to 17 units. It is now making a smaller total profit than before, but it is the largest profit possible now that

1 In the numerical example which has just been discussed, the firm could not select an output which made marginal cost exactly equal to the price of the product (£16), because the marginal cost was £15 at an output of 7 units and £17 at 8 units. This is, however, merely the consequence of the implied assumption that the firm could not 'split' units when selling them. If it could sell half a unit for £8, and if the cost of producing an additional half-unit to bring total output up to 17½ units were £7.75, it would be worth producing the extra half-unit; and it would be worth seeing whether an extra quarter-unit could then be added at a cost of less than £4. The process will stop only when marginal cost is at last equal to marginal revenue for some very small fraction of a unit.

the market price has fallen to P_2 . In the same way, if a further fall in the market price were to reduce it to £18 – indicated by the level of P_3 – the firm will cut output down to 15 units to restore the equality of marginal revenue and marginal cost.

Once the price has fallen to the P_3 level, it is not only equal to marginal cost at an output of 15 units; it is equal to average cost as well, since the point 'C' is where the MC curve intersects the AC curve. The firm's profit is now at the lowest level needed to induce it to continue in business as a producer of this particular commodity; it is the 'minimum' or 'normal' profit which was earlier classified, along with 'capital rental', as one of the costs of production. If the market price were to fall any further, the firm would not even be making this minimum amount, and would eventually withdraw from the market if the price did not subsequently rise again. It might continue in production for the time being (provided the price had not fallen below its average *variable cost*)¹ in the hope of better things; but if the market price continued below P_3 for any length of time, the firm would either switch to another market by producing (if it could) a different commodity, or else go out of business altogether.

Thus, the part of the MC curve beginning at 'C', and running upwards to the right through points 'B' and 'A' and beyond, is analogous to the supply curve as far as this particular firm is concerned; it shows how much of the commodity the firm will want to put on the market at each possible price. However, the supply curve in the market as a whole shows the combined reactions of *all* firms to the various prices they may be offered; so the next step is to see how the market supply curve itself is derived.

TEST 3

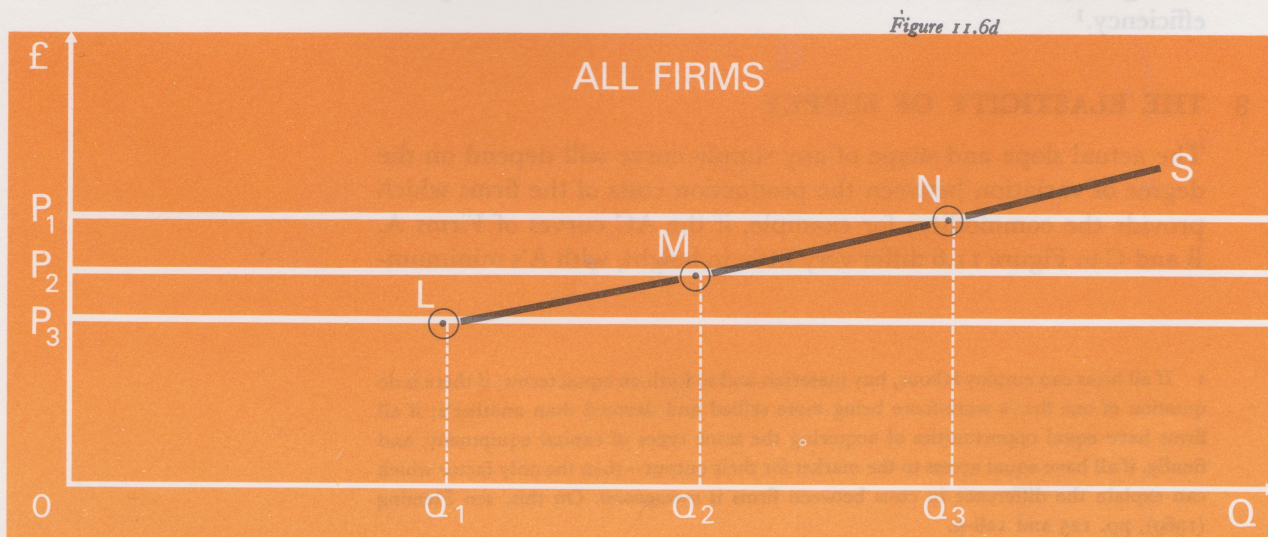
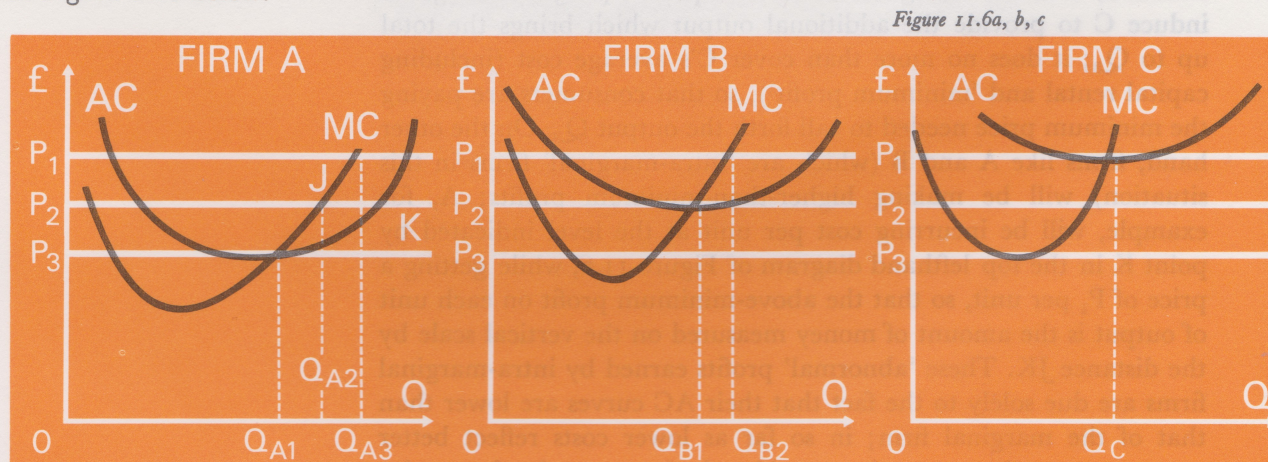
7 'PERFECT COMPETITION' AND THE SUPPLY CURVE

Throughout the preceding discussion, it has been assumed that the individual firm can sell any amount it likes at the price which rules in the market as a whole. For this assumption to be realistic, it is necessary that the firm should be one of a large number, all producing exactly the same commodity; between them, they are producing the quantity consumers wish to buy at the ruling price. Because there is no difference between the individual firms' products, it is not possible for any one of them to charge a higher price than its competitors are asking – if it did, its customers would buy from other firms and it would sell nothing. If it were to ask a *lower* price than its competitors, it would certainly attract more demand; but this would be pointless when it can sell its profit-maximising output (i.e. that at which marginal cost equals price) at the price all other firms are getting. Because the firm's output is only a very small fraction of the turnover in the market as a whole,

¹ If the market price has fallen below the minimum level of average cost, the firm cannot help making a loss for the time being. But if it ceases production altogether, this does not mean that its production costs fall to zero; it must still meet its *fixed* costs, even if it is not using the equipment, building, etc. which they represent. So if there is some level of output at which the price is more than enough to cover the *variable* cost per unit (i.e. the cost of materials, labour, etc.), it will pay to stay in production, since the excess of total revenue over total variable cost will be available as a partial offset to the fixed costs which the firm cannot escape.

it is in no position to influence the market price by varying its own level of output; even if it were to close down altogether, its elimination from the market would have a negligible effect. Nor, indeed, could a large group of firms attempt to manipulate the price by concerted action; if they did, new ones would come into the market and undercut them – provided, of course, that there are no obstacles to prevent new firms being set up, and that everyone knows enough about current prices, costs and profits to be aware of the opportunities the situation offers.

Where all these conditions are satisfied (i.e. many firms, a homogeneous product, easy entry for new enterprises, and full knowledge on everyone's part of the market facts) the situation is one of *perfect competition*. In such a market, the total supply at any given price will be the sum of the outputs of all firms which find it worthwhile to produce the commodity at that price. For example, if there were a hundred firms whose cost-curves were exactly like those shown in Figure 11.5, a market price of P_3 would cause 1,500 units to be put on the market each day. At higher prices, more of the commodity would be supplied; not only would the existing firms be ready to produce more (the firm in Figure 11.5 will raise production from 15 to 17 units a day if the price goes up from P_3 to P_2), but new firms may now enter the market – firms which could not have made even the minimum profit at a price of P_3 , but which can do so at P_2 or P_1 . These are firms whose production costs are higher than those shown in Figure 11.5; the 'minimum cost' points on their AC curves (equivalent to point 'C' in Figure 11.5) are at higher levels, as shown in Figure 11.6 below.



Here, when the market price is P_3 , Firm A will produce Q_{A1} units, and so will every other firm whose cost curves are identical with A's; their combined output is shown as Q_1 in the lower diagram. When the price rises to P_2 , it becomes profitable for Firm B to enter the market, since the price is now just enough to cover its average cost of production at an output of Q_{B1} ; firms with AC curves at levels *between* Firm A's and Firm B's will also begin producing, and Firm A itself will expand output to Q_{A2} , this being the point at which its marginal cost equals the price P_2 . The combined production of all firms will now be the Q_2 shown in the lower diagram. A further price-increase, this time to P_1 , will make it profitable for still more firms to start up, the limiting case being Firm C, which can now just cover its costs at an output of Q_C ; firms whose AC curves are at levels between B's and C's will also be drawn into the market, while A will increase production to Q_{A3} and B to Q_{B2} . Total output from all firms will now be Q_3 in the lower diagram. The points L, M and N in the 'All Firms' diagram show the total supply forthcoming at the three prices. If the same procedure were followed with prices *between* P_1 and P_2 , and P_2 and P_3 , it would provide a set of additional points intermediate between L and N, and a line passing through all of them is the *supply curve* of the market as a whole (marked 'S' in the lower diagram of Figure 11.6).

If the market *demand* curve happens to pass through point N, indicating that consumers wish to buy Q_3 units of the commodity when the price is P_1 , an *equilibrium* will exist in the market when firms have adjusted their combined output to the Q_3 level. In this situation, C is the 'marginal firm'. The price P_1 is just enough to induce C to provide the additional output which brings the total up to Q_3 ; it does no more than cover C's average cost (including capital rental and 'minimum profit'), so that consumers are paying the minimum price needed to call forth the output Q_3 . On the other hand, firms like A and B (which are 'intra-marginal' firms in this situation) will be making higher-than-minimum profits; A, for example, will be incurring cost per unit at the level indicated by point K in the top lefthand diagram of Figure 11.6, while getting a price of P_1 per unit, so that the above-minimum profit on each unit of output is the amount of money measured on the vertical scale by the distance JK. These 'abnormal' profits earned by intra-marginal firms are due solely to the fact that their AC curves are lower than that of the marginal firm; in so far as lower costs reflect better management, they may be considered the reward of superior efficiency.¹

8 THE ELASTICITY OF SUPPLY

The actual slope and shape of any supply curve will depend on the degree of variation between the production costs of the firms which provide the commodity; for example, if the AC curves of Firms A, B and C in Figure 11.6 differ very little in height, with A's minimum-

1 If all firms can employ labour, buy materials and so forth on equal terms; if there is no question of one firm's work-force being more skilled and devoted than another's; if all firms have equal opportunities of acquiring the same types of capital equipment; and finally, if all have equal access to the market for their output – then the only factor which can explain the difference in costs between firms is *management*. On this, see Fleming (1969), pp. 125 and 128–9.

cost point only slightly below P_2 and C 's slightly above it, the market supply curve in the lower diagram will have a very gentle slope. Such a curve is usually described as 'elastic', meaning that a relatively small change in the market price will bring about a relatively large change in the quantity firms are prepared to put on the market.

The 'elasticity of supply' (conventionally denoted by the symbol ϵ_s) is a concept analogous to the 'elasticity of demand' defined in Unit 10.¹ It is the proportionate change in the quantity supplied, divided by the proportionate change in price which gave rise to it.

For example, if the quantity supplied increases 3 per cent when the price rises 1 per cent, ϵ is 3; if quantity increases by 2 per cent when the price rises by 4 per cent, ϵ is $\frac{1}{2}$ – and so on.² If ϵ is 'unitary' (i.e. equal to 1.0), the quantity supplied will be changing in exactly the same proportion as the price; a supply curve whose elasticity is unitary along its whole length will be a straight line passing through the origin ('O') of the diagram, like S_1 in Figure 11.7 below. A line which cuts the vertical scale *above* the origin, like S_2 , has elasticity greater than 1.0, while a line cutting the horizontal scale to the *right* of the origin, like S_3 , has elasticity less than 1.0.

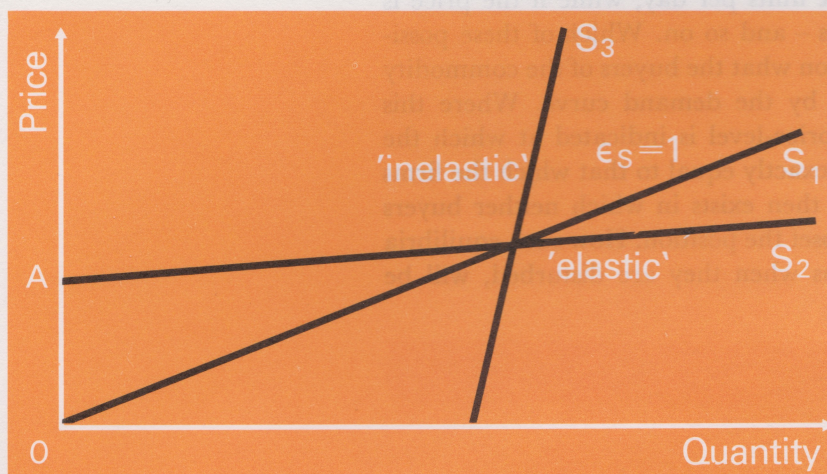


Figure 11.7

There is, of course, no reason why supply curves should necessarily be straight rather than curving. Thus, a typical supply curve might begin at point A on S_2 , run along S_2 itself for a little way (during which its elasticity would be greater than 1.0), then gradually become steeper until its slope approximates to that of S_1 (so that the $\epsilon_s = 1$), then become steeper still so that it runs along S_3 (and is thus 'inelastic') and finally become vertical – at which point its elasticity would be zero.

It has already been noted (on p. 46 above) that the shape of firms' cost curves is likely to depend to a considerable extent on the time-perspective which is adopted. The same thing applies to market supply curves. Over a very short period of time such as a single day, firms may be quite unable to alter their levels of output, whatever may be happening to the prices of their products; it will also be out of the question for new firms to enter the market at such short notice. So *very* 'short-run' supply curves are highly inelastic, approxi-

¹ See Unit 10, page 24.

² M 3.

mating to vertical straight lines. Over the rather longer 'short-period' (as defined on p. 46 above) during which firms are unable to make significant changes in the amount and nature of their capital equipment, some adjustment of the quantity supplied will be possible; established firms will have the opportunity to change their output, and new firms will have time to enter the market; this will give a rather more elastic supply curve of the kind derived in the lower diagram of Figure 11.6. In the still longer run, it will be possible for firms to alter and replace equipment and to adopt new techniques of production, so that supply will become still more 'elastic'. Thus, the response of output and supply to price-changes will normally be greater, the longer the time-period within which adjustment is possible.

9 SUPPLY, DEMAND AND THE MARKET

Whatever the amounts which firms are prepared to supply at particular prices, it is necessary that they should find buyers for their commodities; so at this point, *demand* must be brought back into the picture. No supply curve can do more than indicate a range of possibilities: if the market price happens to be xp per unit, firms will be prepared to sell X units per day, while if the price is yp they will want to sell Y units – and so on. Which of these possibilities will materialise depends on what the buyers of the commodity are prepared to do, as shown by the demand curve. Where this intersects the supply curve, a price-level is indicated at which the quantity buyers wish to have is exactly equal to that which the firms wish to sell, and an *equilibrium* then exists in which neither buyers nor sellers have any reason to upset the position. How such equilibria are reached, and what happens when they are disturbed, will be considered in detail in Unit 12.

TEST 4

MATHEMATICAL NOTES

- M** 1 Like demand curves, supply curves can also be represented by equations. An example is

$$S = 5P - 100$$

where S is the number of units of a commodity that sellers are willing to sell per day and P is the price in pence. If the price of the commodity happened to be 25p, the equation says that sellers would be prepared to put 25 units on the market per day – i.e. $(5 \times 25) - 100$. If the price were 30p, it can be seen by putting 30 in place of P in the equation that the quantity supplied would be 50 units per day. In the same way, a price of 50p would cause supply to be 150 units a day. At a price of 20p, nothing would be supplied at all, and the equation must be regarded as having no meaning at prices *below* 20p (since it would then indicate that a *negative* quantity was being offered).

This equation, if plotted on a diagram, would give a straight line. But there is no reason why supply curves should not be *curved*, and the appropriate equation would then be more complicated. The *general* functional relationship can be indicated by

$$S = S(P)$$

where the $S(\dots)$ stands for the 'supply function'. It should be noted that this expression implies that P is an 'independent variable' – that is, it implicitly assumes that sellers cannot themselves deliberately control the price-level, but merely decide how much to offer for sale in view of the price which is ruling in the market.

- M** 2 Thus, marginal cost is the derivative of *total* cost with respect to output, i.e.

$$MC = \frac{dC}{dQ}$$

where C is the total cost and Q is the quantity produced.

- M** 3 More generally, the price-elasticity of supply can be expressed as

$$\frac{\Delta Q_s}{Q_s} \cdot \frac{P}{P}$$

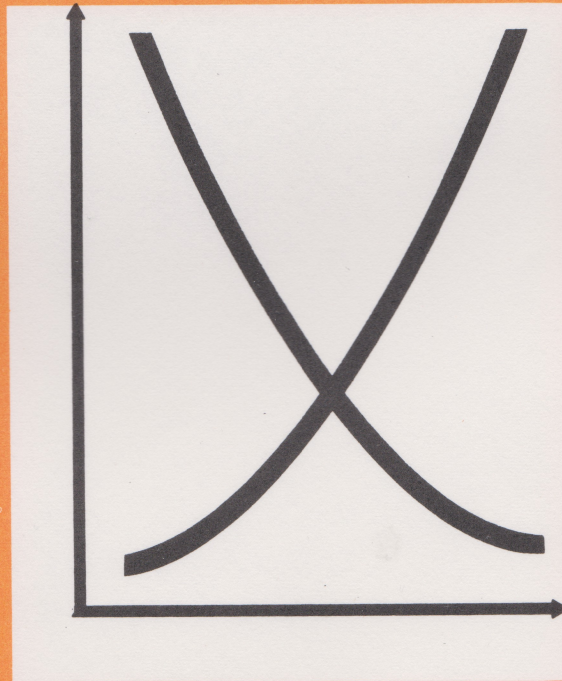
where Q_s is the quantity supplied, P is the market price, and ' Δ ' denotes a small change in the magnitude it precedes. This gives

$$\epsilon_s = \frac{\Delta Q_s}{\Delta P} \cdot \frac{P}{Q_s}$$

- FLEMING, J. M. (1969). *Introduction to Economic Analysis*. London, Allen and Unwin.
- MARSHALL, ALFRED (1961). *Principles of Economics*, 9th edition. London, Macmillan.
- SAMUELSON, P. A. (1970). *Economics: an Introductory Analysis*, 8th edition. New York, McGraw-Hill.

Unit 12

Markets and Prices



CONTENTS UNIT 12

	PAGE
1 SUPPLY AND DEMAND	61
2 SHIFTS IN SUPPLY AND DEMAND	63
2.1 Shifts in the Supply Curve	64
2.2 Shifts in the Demand Curve	67
2.3 Movement from One Equilibrium to Another	69
3 MONOPOLY	69
3.1 Monopolistic Demand Curve	71
3.2 Profit-Maximising for a Monopolist	73
3.3 Monopolistic Competition	75
3.4 Profit-Maximisation with Monopolistic Com- petition	75
3.5 Oligopoly	78
4 THE FUNCTIONS OF THE PRICE SYSTEM	79
4.1 Allocation of Resources	79
4.2 Efficiency of Allocation	80
4.3 The Price System and Technical Progress	82
4.4 Limitations of the Price System	82
5 FACTORS OF PRODUCTION AND THE PRICE SYSTEM	85
6 CONCLUSION	85
BIBLIOGRAPHY	86

MARKETS AND PRICES

1 SUPPLY AND DEMAND

Unit 10 introduced the demand curve and explained why it usually slopes downwards to the right. Unit 11 introduced the supply curve and explained why it slopes upwards to the right. The demand curve indicates the quantities of a commodity consumers wish to purchase at each particular price, while the supply curve indicates the quantities producers wish to sell at each particular price. The point at which the demand curve intersects the supply curve indicates the price at which the amount consumers wish to buy equals the amount producers wish to sell. This price is known as the *equilibrium price* because both buyers and sellers are satisfied with the amount bought and sold and the price charged, and there is therefore no reason to expect price to change.

This situation is shown in Figure 12.1 which combines the demand curve introduced in Unit 10¹ and the supply curve introduced in Unit 11². The point P, at which the demand curve DD intersects the supply curve SS, is where the price is 20 pence and the quantity demanded and offered for sale is 150,000 units per day.³

Consider the situation if the price was 15 pence. Looking at Figure 12.1 you can see that at this price producers are willing to offer for sale 100,000 units per day, whereas buyers are demanding some-

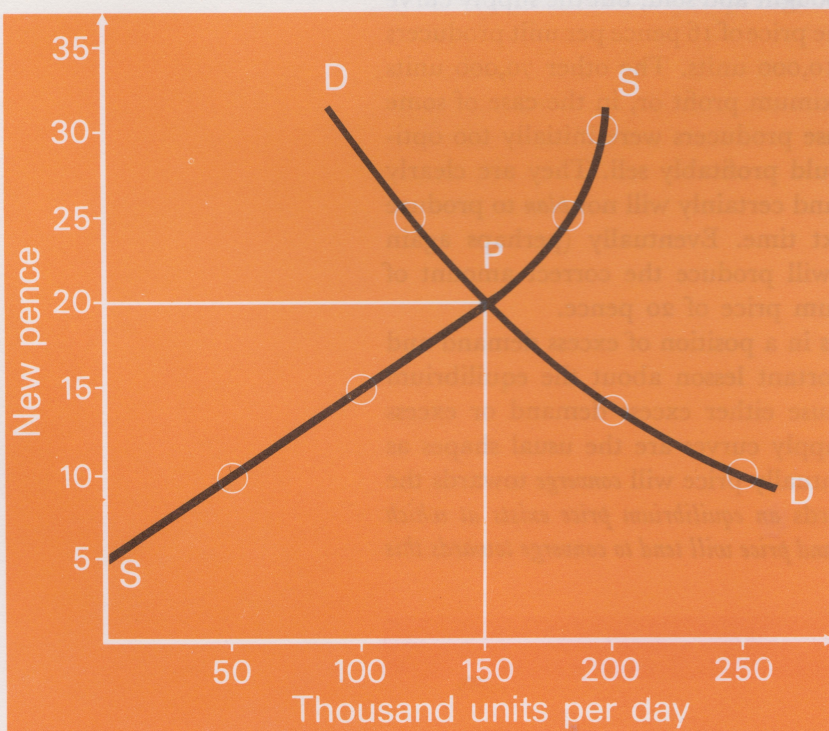


Figure 12.1

1 See Unit 10, Figure 10.1, p. 11.

2 See Unit 11, Figure 11.1, p. 36.

3 Remember the explanation given in Units 10 and 11 that the supply and demand curves refer to a set of simultaneous alternative possibilities and not to the changing situation in different time-periods.

thing close on 200,000 units per day. This is called a situation of *excess demand*. Sellers are satisfied that they are selling all they wish to sell. Buyers are clearly dissatisfied at only being able to obtain about half as much as they wanted.

In this situation sellers have an incentive, in the form of larger profits, to raise the price and offer more for sale in accordance with their supply curve. At first they may still not produce enough and there will still be excess demand. Or they may produce too much and find that they can't sell it all at the price they wish to charge. Eventually (even if only by a process of trial and error), the equilibrium price will be established at which the amount supplied equals the amount demanded.

This process occurs in reverse if the initial price is higher than the equilibrium price. For example, if the price was 25 pence the amount demanded would be about 120,000 per day and the amount supplied about 180,000 per day.¹ Firms will find large surpluses of stocks on their hands which they are unable to sell. This is known as a situation of *excess supply*. Buyers are satisfied in being able to obtain all they demand but sellers are clearly dissatisfied in not being able to sell all they have produced. They will have to reduce the price to sell their surplus stock. In fact to induce buyers to take all the 180,000 units they have produced, they will have to reduce price below the equilibrium level. This is because at the equilibrium price of 20 pence only 150,000 units are demanded and the price will have to fall further than this to sell the other 30,000. As Figure 12.1 shows, the price at which 180,000 are demanded is about 16 pence.

This, however, is not the *equilibrium* price. It is true that at 16 pence 180,000 units have been bought and sold, but the supply curve in Figure 12.1 indicates that at the price of 16 pence per unit producers only wished to supply about 110,000 units. The other 70,000 units have been sold at less than maximum profit or, in the case of some firms, probably at a loss, because producers were initially too optimistic as to how much they could profitably sell. They are clearly not satisfied with this situation and certainly will not *plan* to produce 180,000 units at 16 pence next time. Eventually (perhaps again through trial and error) they will produce the correct amount of 150,000 to sell at the equilibrium price of 20 pence.

This analysis of what happens in a position of excess demand and excess supply provides an important lesson about the equilibrium price. If price is such as to cause either excess demand or excess supply and the demand and supply curves are the usual shapes as shown in Figure 12.1, then eventually price will *converge* towards the equilibrium price. In other words *an equilibrium price exists at which both buyers and sellers are satisfied, and price will tend to converge towards this equilibrium level.*²

TEST 1

¹ See Figure 12.1.

² There are cases where price may *diverge* from the equilibrium level. This occurs where either the supply or demand curve is not the usual shape. In this situation if price is not at the equilibrium level the reactions of buyers and sellers will only cause a further divergence. This, however, is a rather special case which can be ignored for the time being.

2 SHIFTS IN SUPPLY AND DEMAND

Supposing an equilibrium price exists as in Figure 12.1. What happens if the supply or demand curve changes? These changes can occur for a variety of reasons. The shape and position of the supply curve depend on the various factors of production, material inputs, and the manner in which they are organised to produce the good. If the price of a factor of production rises (for example, a wage rise occurs) this might raise the price of the product or alternatively reduce the amount sellers are willing to offer for sale at each given price. In Figure 12.1 sellers were willing to offer 150,000 units for sale at a price of 20 pence. If wages rise, then some sellers will no longer want to produce the good at a price of 20 pence and so the overall supply will fall. Alternatively there may be some technological breakthrough which reduces the cost of the good and enables more to be sold at any given price. These two situations are shown in Figure 12.2.

The original situation is shown by the supply curve SS_1 . If there is a wage rise some firms can no longer produce the good profitably at a price of 20 pence, so supply is reduced and the supply curve shifts to the left, to curve SS_2 . And if there is some technological advancement and the cost is reduced, more can be produced at 20 pence, supply is increased and the supply curve shifts to the right to SS_3 . When supply increases or decreases and the supply curve shifts either to the left or right this situation is known as a change in the *conditions of supply*.¹ In other words one of the other variables besides price in the supply function has changed.

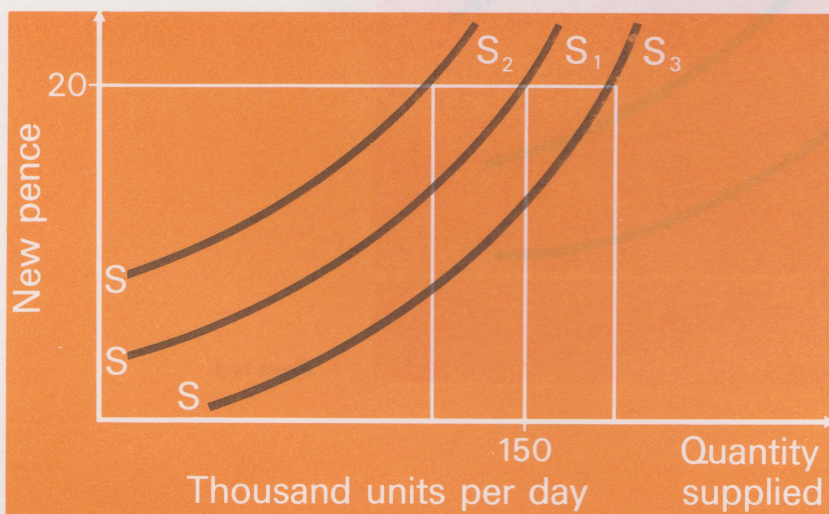


Figure 12.2

¹ The student should be sure he understands the difference between 'an increase in the quantity supplied' and 'an increase in supply'. A supply curve measures only the effect of price on the quantity supplied. It assumes that all other factors such as the cost of the factors of production and the manner in which they are organised are fixed. Often, however, as has just been explained in the main text, these factors (for example, wage rates) will change. The supply curve which assumed them to be fixed is no longer operational and a new supply curve comes into operation. Whenever a change in the other factors affecting the supply curve causes more or less to be supplied at any given price this is known as an *increase or decrease in supply*. Whenever the other factors are constant and a change in price causes more or less to be supplied this is called an *increase or decrease in the quantity supplied*. In terms of graphical presentation the former denotes a shift in the supply curve either to the right or left, whereas the latter denotes a movement up or down a given supply curve.

The shape and position of the demand curve depends on such things as people's incomes, their wealth and tastes and the prices of other goods which might be complements or substitutes for the good in question.¹ If any of these change, then the demand curve must change. For example, an increase in income may mean that some consumers now consider buying a particular product which they previously thought too expensive. So more will be bought at any given price, demand is increased and the demand curve will shift to the right. Alternatively, the price of a substitute good may fall, so that a number of people switch to buying the substitute and less is demanded of the original good at any given price. These two situations are shown in Figure 12.3 where the original demand curve is DD_1 as in Figure 12.1. With this demand curve 150,000 units were demanded at 20 pence. An increase in income causes demand to increase, so more is demanded at the price of 20 pence and the demand curve shifts to the right to DD_2 . A fall in the price of a substitute causes demand to decrease, so less is demanded at the price of 20 pence and the demand curve shifts to the left to DD_3 . A *shift* in the demand curve indicates an increase or decrease in demand due to a change in the *conditions* of demand. In other words one of the variables in the demand function has changed.²

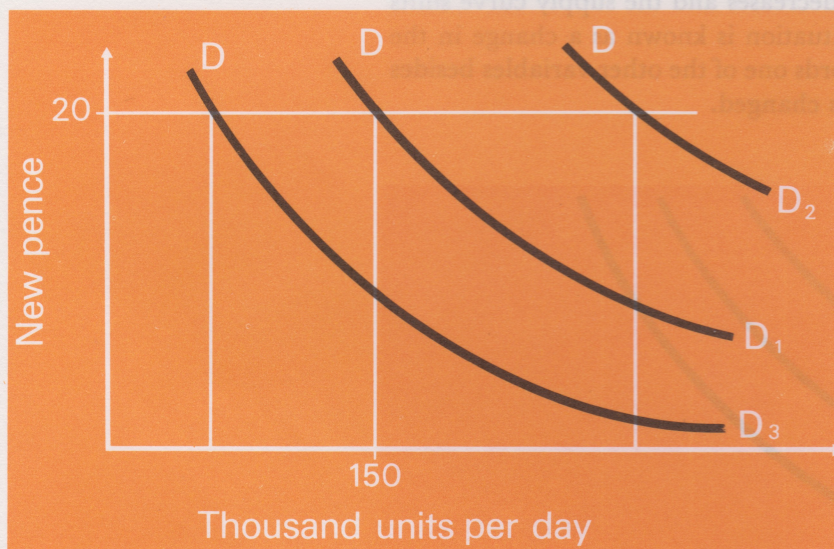


Figure 12.3

2.1 Shifts in the Supply Curve

To return to the original question, what happens to the equilibrium price if the supply or demand curve changes? Suppose supply increases for the reason given earlier and the supply curve shifts to the right. The result is shown in Figure 12.4 below.³

The original supply curve SS_1 cuts the demand curve DD at point X to create an equilibrium price of P_1 and an equilibrium output of Q_1 . The supply curve now shifts to the right to SS_2 indicating that more will be supplied at any given price. The new

¹ See Unit 10, pp. 20–22.

² The footnote on p. 63 explaining the difference between 'an increase in the quantity supplied' and 'an increase in supply' also applies to demand.

³ The supply and demand curves are drawn as straight lines to simplify the exposition.

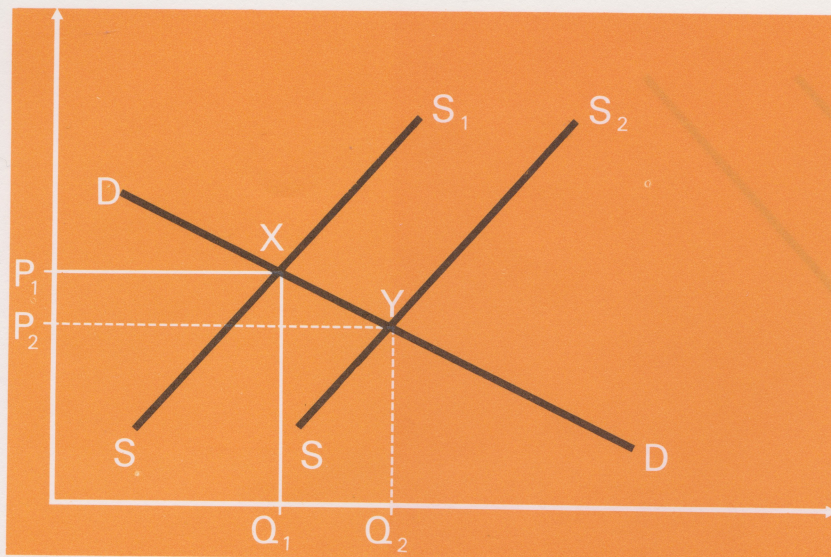


Figure 12.4

supply curve cuts the demand curve¹ at point Y, creating the new equilibrium price P_2 and the new equilibrium output Q_2 . The effect of the supply curve shifting to the right is that more is sold and at a lower price than previously.

This, however, is not very precise. In Figure 12.4 both price and output are affected by the shift in the supply curve, but clearly the largest effect has been on output with price having fallen by only a relatively small amount. *Whether price or output will be relatively more affected by a shift in the supply curve will depend on the elasticity of demand.*² In Figure 12.4 the demand curve is elastic³ indicating that the price only has to change by a small proportion to induce consumers to buy more. So when the supply curve shifts to the right, indicating that suppliers are willing to supply more at any price, the price only has to drop by a relatively small proportion to induce consumers to buy the extra output.

Figure 12.5 indicates the effect if demand is inelastic. In this situation price has to fall by a relatively large proportion before consumers are persuaded to buy more units of the good.

The new supply curve S_2 creates a new equilibrium price P_2 , which is a substantial proportion below the old equilibrium price P_1 , and a new equilibrium output Q_2 , which is only a small proportion above the old equilibrium output Q_1 . This should be contrasted with Figure 12.4 where the effect was exactly the opposite. The student should pause for a moment and consider what would happen to price and output if demand was completely inelastic (i.e. a vertical line) or completely elastic (i.e. a horizontal line). It should be clear that a completely inelastic demand curve (indicating that no amount of reduction in the price would induce consumers to buy more) means that an increase in supply results in a relatively substantial fall in price and no increase in output. Conversely a completely elastic demand curve (indicating that consumers will buy as much as they can at any given price) means that an increase in supply results

¹ There is no reason to assume that the demand curve will change when the supply curve changes. There are certain circumstances when it might but to make the exposition simpler the demand curve is assumed not to change.

² To refresh your memory of this concept refer to Unit 10, pp. 24–5.

³ I.e. the percentage change in output is larger than the percentage change in price.

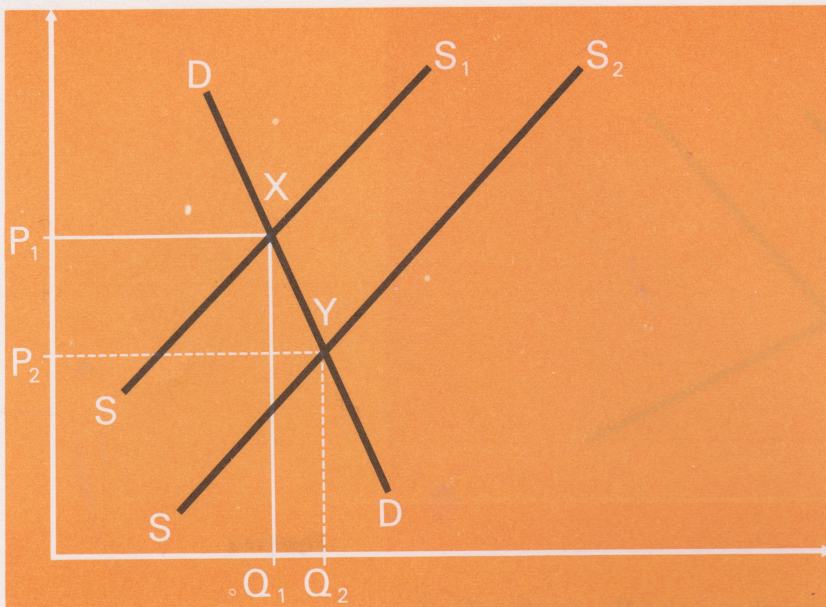
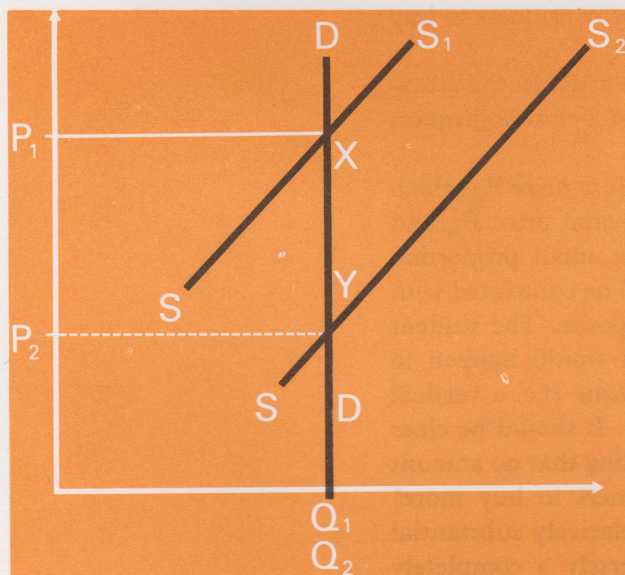


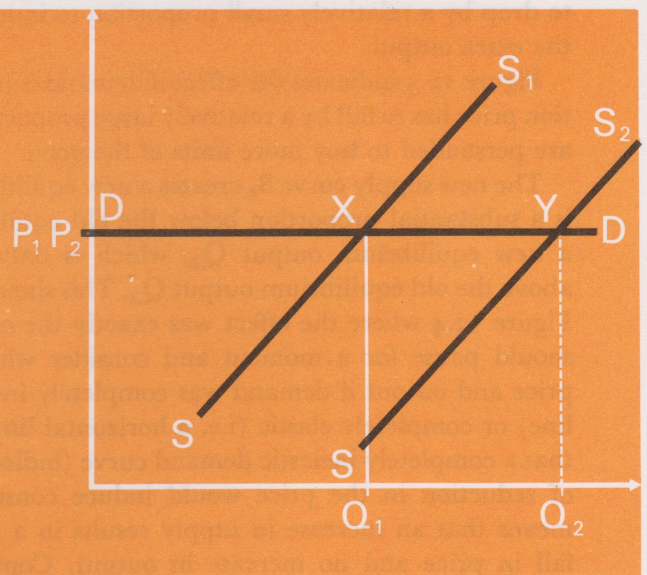
Figure 12.5

in no price reduction and a relatively substantial output increase. These two cases are shown in Figures 12.6 and 12.7.

The completely inelastic demand curve produces a relatively large fall in price and no rise in output. Therefore **the more inelastic is the demand curve the greater will be the relative effect on price rather than on output of an increase in supply.** Conversely the completely elastic demand curve produces a relatively large rise in output and no fall in price. Therefore **the more elastic is the demand curve the greater will be the relative effect on output, rather than on price, of an increase in supply.**



Completely Inelastic Demand Curve



Completely Elastic Demand Curve

Figures 12.6 and 12.7

Producers may also be willing to supply less at any given price than previously, and this is shown by a shift of the supply curve to the left. The effect of this on the equilibrium price and output is shown in Figure 12.8.

The effect is the exact opposite of what happens if supply increases and the supply curve shifts to the right as shown in Figure 12.4. Now

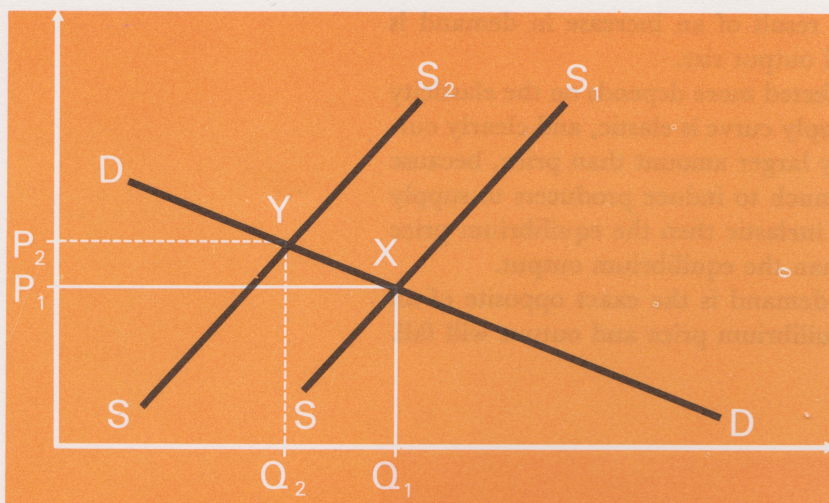


Figure 12.8

SS_1 is the original supply curve intersecting the demand curve at X and producing an equilibrium price and output of P_1 and Q_1 . The supply curve shifts to SS_2 giving a higher equilibrium price P_2 and a lower equilibrium output Q_2 . So when the suppliers offer less for sale at any given price than previously the result is that price rises and output falls.

Whether price or output will be more affected by the reduction in supply depends, as has just been explained, on the elasticity of demand. In Figure 12.8 demand is elastic so that output is relatively more affected than price. If demand was inelastic price would be relatively more affected than output.

2.2 Shifts in the Demand Curve

What happens if consumers are willing to purchase more at any given price? The demand curve will shift to the right – the result is shown in Figure 12.9. The original demand curve DD_1 cuts the supply curve SS at X, giving an equilibrium price P_1 and an equilibrium output Q_1 . The new demand curve DD_2 cuts the supply curve¹ at Y, giving the new equilibrium price P_2 and the new

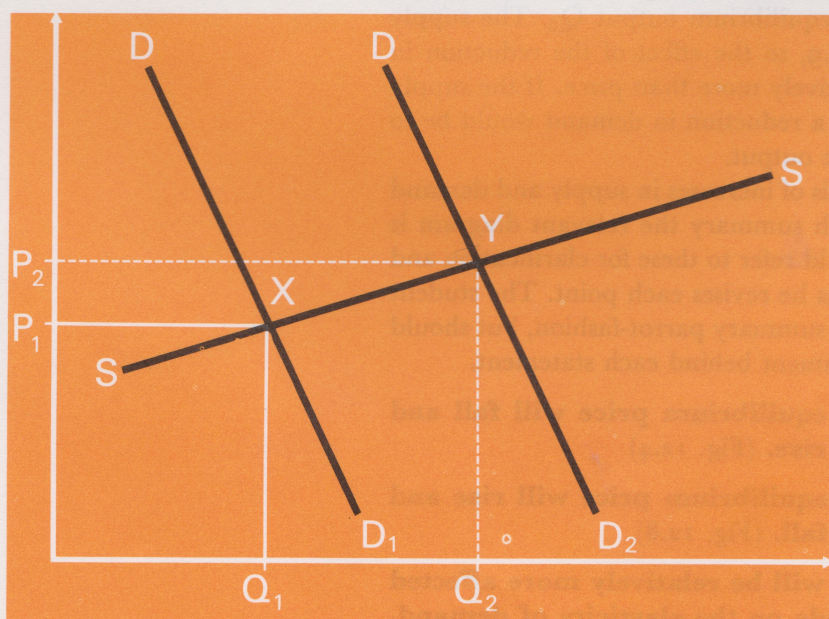


Figure 12.9

¹ The supply curve is assumed not to change as the demand curve changes.

equilibrium output Q_2 . So the result of an increase in demand is that both equilibrium price and output rise.

Whether price or output is affected more depends on the elasticity of supply. In Figure 12.9 the supply curve is elastic, and clearly output has increased by a relatively larger amount than price, because price does not have to rise by much to induce producers to supply more. If the supply curve were inelastic then the equilibrium price would rise by relatively more than the equilibrium output.

The effect of a reduction in demand is the exact opposite of an increase in demand, namely equilibrium price and output will fall. This is shown in Figure 12.10.

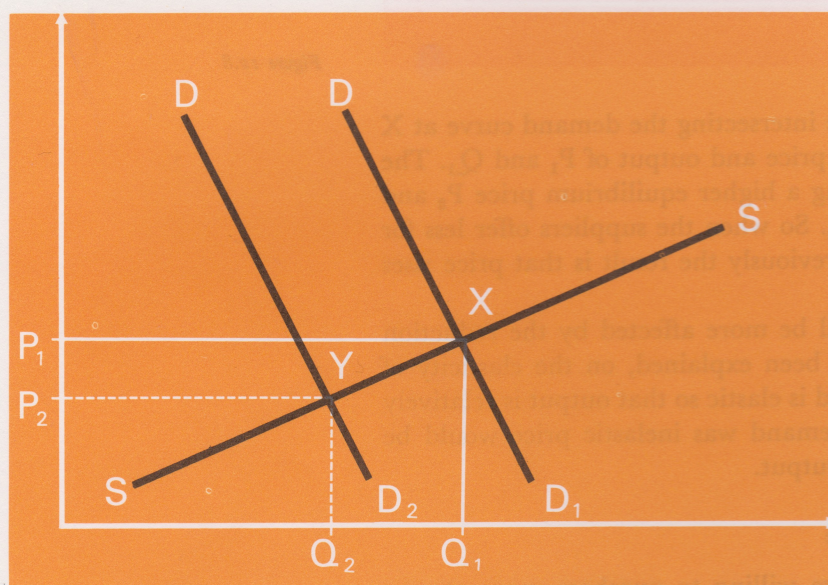


Figure 12.10

In Figure 12.10 the original demand curve DD_1 cuts the supply curve at X, producing the equilibrium price P_1 and the equilibrium output Q_1 . A reduction in demand causes a shift of the demand curve to DD_2 producing an intersection with the supply curve at Y and a new equilibrium price P_2 and equilibrium output Q_2 . The supply curve is elastic as in Figure 12.9, so the effect of the reduction in demand is to lower output relatively more than price. If the supply curve was inelastic the effect of a reduction in demand would be to reduce the price rather than the output.

The main points of the analysis of increases in supply and demand are now summarised. After each summary the relevant diagram is referred to and the student should refer to these for clarification and attempt to draw them himself as he revises each point. The student should not attempt to learn this summary parrot-fashion, but should be sure he understands the argument behind each statement.

- 1 If supply increases, the equilibrium price will fall and the equilibrium output will rise. (Fig. 12.4)
- 2 If supply decreases, the equilibrium price will rise and the equilibrium output will fall. (Fig. 12.8)
- 3 Whether price or output will be relatively more affected by a change in supply depends on the elasticity of demand. The more inelastic is demand the greater will be the relative effect on price rather than output, and vice versa the more elastic is demand. (Figs. 12.4, 5, 6 and 7)

4 If demand increases both the equilibrium price and the equilibrium output will rise. (Fig. 12.9)

5 If demand decreases both the equilibrium price and the equilibrium output will fall. (Fig. 12.10)

6 Whether price or output will be relatively more affected by the change in demand depends on the elasticity of supply. The more inelastic is supply the greater will be the relative effect on price rather than on output, and vice versa the more elastic is supply. (Figs. 12.9 and 10 and supporting text)

2.3 Movement from One Equilibrium to Another

The analysis has so far not indicated how price will move from one equilibrium to another. It is clear from the previous analysis of how the initial equilibrium price was found that this does not occur automatically. This is a complex matter and full discussion must await a more advanced treatment. Briefly it may be said to depend mainly on how sellers react to a change in the conditions of supply and demand.

Suppose, for example, that consumers wish to buy more at any given price and their demand curve shifts to the right. Suppliers may not notice this until either their stocks begin to run out at a much faster rate or else perhaps certain customers offer a higher price in order to obtain the product. In many cases suppliers will not be able to adjust their output very quickly to the new situation. Their capital will be fixed and extra labour may be difficult to obtain. In this event the increase in demand will not bring forth much extra output in the short run. The supply curve will, in effect, be very inelastic and the effect of the increase in demand will be to raise the price until the demand is reduced to meet the limited supply.

In the longer run firms can install new equipment, take on more labour, etc. and expand their output. As more output becomes available to satisfy the increased demand, the market price will be reduced again – though not all the way back to its original level. Given time for the adjustment of production, the supply curve is more elastic than it is in the very short run.

It has been argued that in fact for many goods the equilibrium price and output levels are never reached. By the time supply has reacted to the new demand situation some further changes have occurred and a new equilibrium level has to be reached. So price and output will continue to move around the equilibrium level without ever reaching it. This takes us into dynamic analysis which requires a more complex approach than is possible here. For the time being it is sufficient for the student to grasp the concept of the *new* equilibrium level of price and output resulting from an increase or decrease in demand or supply.

TEST 2

3 MONOPOLY

The analysis in the last unit and, so far, in this, has assumed that the supply curve sums up the reactions of a great many firms each with a small share of the total market. This may not always (or often) be the case. Sometimes the total supply will be provided by just one firm, a situation known as *monopoly*. How will this affect the equi-

Equilibrium price and output? Before answering this question it is important to be clear about what exactly a monopoly is.

It has been assumed up to now that the market for a particular good is supplied by a large number of firms and consumers see no difference between individual firms' products. Because of this a firm cannot raise the price of its own product because customers will switch to its competitors. And there is no incentive to lower the price when it can sell its profit-maximising output at the price all other firms are getting.¹ So the firm takes the price as given, and its demand curve is a horizontal straight line, i.e. it produces as much as it finds profitable at the market price.

In fact, of course, it is very rare for there to be no difference in the eyes of consumers between individual firms' products in the same market. When consumers do not regard products as complete substitutes for one another, an element of monopoly occurs. A firm can then raise the price of its product and know that some consumers will not switch to an alternative product because they do not regard it as a complete substitute.

Sometimes this element or degree of monopoly will be very small; that is when consumers still regard two products as fairly close, but not quite complete, substitutes. Jaffa oranges, for example, may be preferred to Spanish oranges. The producers of Jaffas may therefore be able to raise their price by a small amount and not lose too much custom. However, because the difference between the two oranges is not very marked, a larger rise in price is likely to cause most people to switch, and so the degree of control that the Jaffa producer has over his price is limited.

In contrast, if one person was able to control the total supply of water he would have a complete monopoly because there is no substitute for this commodity. He could charge a very high price and still be sure of having a demand for his product. His demand curve would be very inelastic.

These two extreme cases indicate that *the degree of substitutability* for the firm's product decides the extent of its monopoly. Often a firm may be the sole supplier of a particular product and not have a very strong monopoly. The National Coal Board is virtually the only producer and supplier of coal in the U.K. However, because there are effective substitutes for coal as a source of heat and power its degree of monopoly is not very strong. The electricity industry on the other hand has a very strong monopoly over the lighting market as no effective substitute exists.

So it is the degree of substitutability which decides the extent of a monopoly. Where there is complete substitutability no firm can individually raise its price and the demand curve for the individual firm is horizontal (perfectly elastic). Where there is no substitute for a particular product a complete monopoly exists, and the demand curve is perfectly inelastic.² In between these two extremes the slope

1 See Unit 11, p. 50.

2 In fact no good has an absolutely complete monopoly as all goods compete in some way for consumers' demand. If the price is too high people may just decide to go without the good. Where goods are essential like water the monopoly is limited by the amount of money (or other assets) that people have. They may be prepared to spend a lot of money to obtain water, but they can't spend more than they own or can borrow. So even a water monopolist is to some extent limited by the price he can charge. Therefore Figure 12.11 (b) is not likely to occur in reality.

of the demand curve, indicating the degree of monopoly, will depend on what substitutes exist. These three situations are shown in Figure 12.11 below.

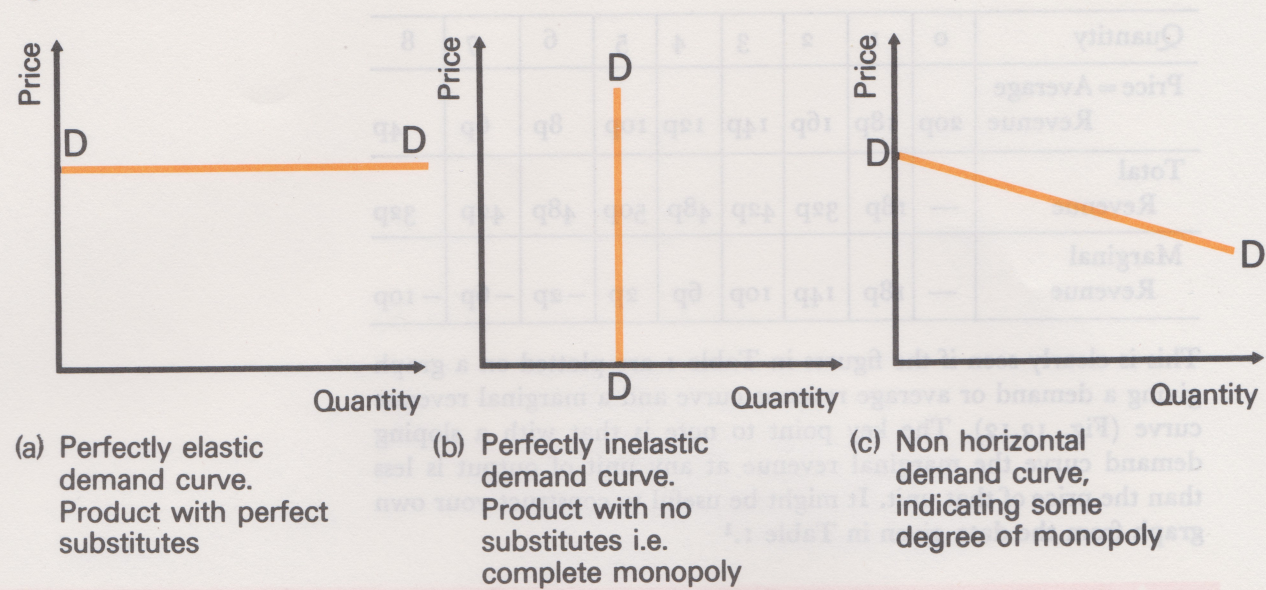


Figure 12.11

3.1 Monopolistic Demand Curve

The above analysis indicates that a firm with some monopoly power over its product will have a sloping demand curve instead of the horizontal one previously assumed in a perfectly competitive market. This means that it has some control over the price of its product. It can increase the price and sell less, or lower the price and sell more. Unlike the firm with a horizontal demand curve, it knows that when it raises its price it may lose some customers but will certainly not lose all of them.

This has important implications for the marginal revenue curve. With a horizontal (perfectly elastic) demand curve each unit of output is sold at the same price and so the extra revenue gained from each sale (marginal revenue) is always the same and equals the price of the product. With a sloping demand curve the firm has to reduce the price of *all* its output to sell any extra units.¹ This can be shown by the example in Table 1.

Table 1 shows a demand schedule for a firm with a sloping demand curve. In order to increase the quantity sold from 3 to 4 units it has to reduce the price of all its units from 14 pence to 12 pence. So in order to sell one extra unit at 12 pence it has to reduce the price of its 3 other units by 2 pence. The extra revenue gained from selling the fourth unit is not the 12 pence it sold it at but only 6 pence. This is shown by the marginal revenue row in the table. It can be seen from this table that if the firm's demand curve slopes downwards to the right, indicating that the price has to fall to sell extra units of output, then marginal revenue will also decline as output increases. In fact with price – which is the same as average revenue – falling, marginal revenue will always be below average revenue. Compare the average and marginal revenue rows in

¹ This assumes he cannot discriminate between customers and charge some the higher price and some the lower price.

Table 1 and this becomes apparent.

Table 1

Quantity	0	1	2	3	4	5	6	7	8
Price = Average Revenue	20p	18p	16p	14p	12p	10p	8p	6p	4p
Total Revenue	—	18p	32p	42p	48p	50p	48p	42p	32p
Marginal Revenue	—	18p	14p	10p	6p	2p	-2p	-6p	-10p

This is clearly seen if the figures in Table 1 are plotted on a graph giving a demand or average revenue curve and a marginal revenue curve (Fig. 12.12). The key point to note is that with a sloping demand curve the marginal revenue at any unit of output is less than the price of that unit. It might be useful to construct your own graph from the data given in Table 1.¹

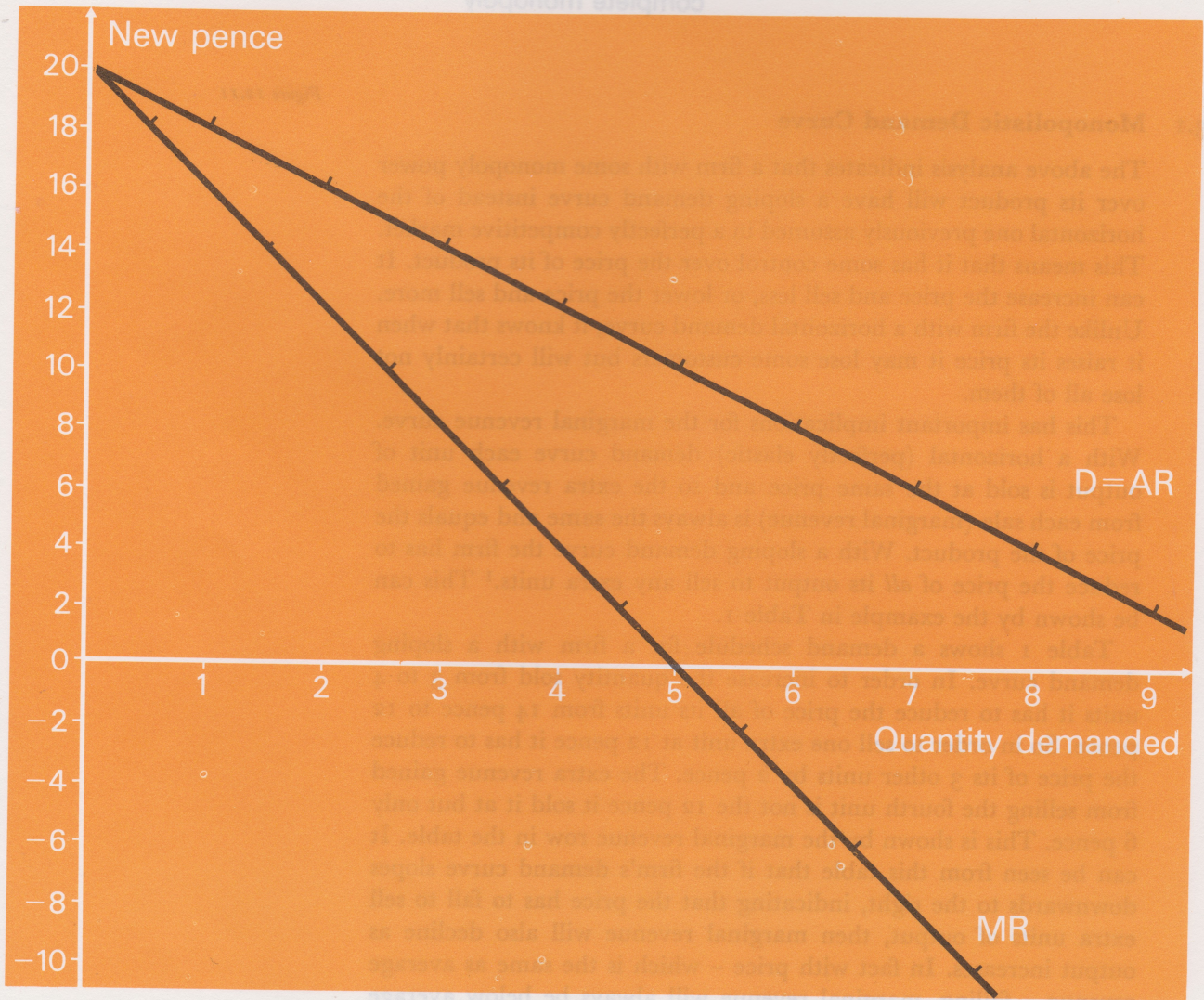


Figure 12.12

¹ If you do this remember that as marginal revenue is created in the process of selling an extra unit the MR curve cuts through points midway between each extra unit of output, e.g. a marginal revenue of 18p is earned by selling the first unit and the line therefore passes through a point midway between the zero and the first unit. Similarly a marginal revenue of 6p is earned by selling the fourth unit and the line therefore passes through a point midway between the third and fourth unit.

While marginal revenue is always less than price it is still positive over a considerable part of the curve (until more than 5 units are sold) which indicates that the firm does obtain some extra revenue from increasing its sales up to that point despite having to reduce its price to do so. This is linked to the *elasticity* concept which was explained in Unit 10.¹ As long as marginal revenue is positive, demand must be elastic (i.e. the percentage increase in quantity demanded is greater than the percentage fall in price) because the firm obtains extra revenue from reducing its price. So the elasticity of demand for its product is an important concept for a firm with a degree of monopoly over its product. It needs to know the effects of raising or lowering its price on its output in order to obtain the maximum revenue.

3.2 Profit-Maximising for a Monopolist

The analysis now returns to the original question. If a firm has some degree of monopoly over its product how will this affect the equilibrium price and output? To answer this question the type of monopoly must be more clearly specified. In the first instance the assumption is that the firm is the only producer of a particular good and no other firms can enter the market to compete with it. Different assumptions will be made later.

The first point that needs to be stressed is that the profit-maximising rule that was introduced under perfect competition in Unit 11 also applies under conditions of monopoly; namely that the firm should choose the output at which marginal cost equals marginal revenue. This profit-maximising rule applies under all types of market situation. For if marginal revenue is above marginal cost the firm is continuing to make profits on its extra output and it would pay it to continue expanding output. And if marginal cost is above marginal revenue the firm is making losses on its extra output and it would pay it to reduce its output.²

In a competitive market, however, marginal revenue equals price so that the firm expands output until *marginal cost equals price*. In a monopolistic market marginal revenue is less than price so that *marginal cost* when it equals marginal revenue *will also be less than price*. This is shown on the graph below (Fig. 12.13). The graph consists of the marginal and average cost curves (which are assumed to be the same for the monopolist as for the competitive firm) and the marginal and average revenue curves which are taken from Figure 12.12. MC refers to the marginal cost curve; MR to the marginal revenue curve; AC to the average cost curve and $D=AR$ to the demand curve which is the same as the average revenue curve. You should by now understand the meaning of these terms and why the curves have the particular shapes shown in Figure 12.13.

It has been stated that profit will be maximised where $MC=MR$, and this is shown to be point X. As Figure 12.13 shows by the dotted line at the point where $MC=MR$, the price is Y, so that MC is well below price, being only Z at that point. This indicates the importance of the elasticity of demand to the monopolist. He must have some idea of the slope of his demand curve if he is to judge the correct

¹ Unit 10, pp. 24-5.

² This argument was more fully explained in Unit 11, pp. 47-50.

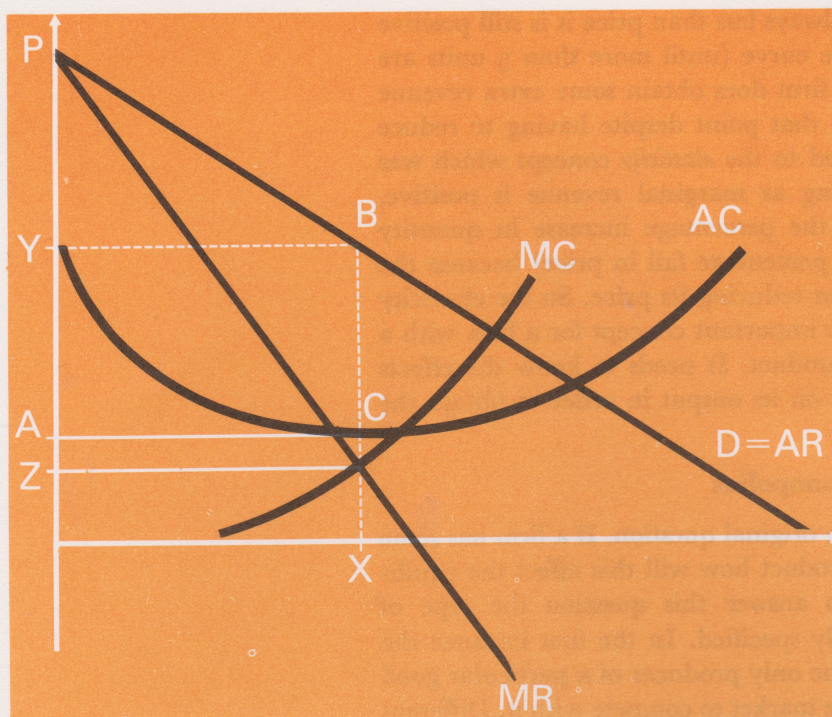


Figure 12.13

output and price figures at which he will maximise profits. The firm in a perfectly competitive market only has to know what its marginal cost is, because it can take the price and therefore the marginal revenue of its product as given. It raises its production levels until marginal cost equals price.¹

The profit the monopolist makes on the unit of output at the point where $MC=MR$ is shown by the difference between his average revenue and his average cost curve.² The total profits he makes at this point are shown by the area YACB in Figure 12.13 which sums together the difference between average revenue and average cost of total output up to this point. Remember that the definition of costs included an element of normal profits, so this area shows the excess profits that the monopolist makes because other firms cannot enter the market to reduce the demand for his particular product and therefore his profits.

The theory of the profit-maximising behaviour of the firm in a monopolistic market may seem complex, but there are a few basic elements to it which will now be repeated. First it builds on concepts which have already been introduced in Unit 11. You need to be sure you understand the following:

The meaning of average cost, marginal cost, average revenue and marginal revenue.

The shape of the curves for these concepts for a firm in a perfectly competitive industry.

The profit-maximising position of a firm in a perfectly competitive industry.

1 As a useful self-test, work out for yourself what the price and output would be if the monopolist produced until marginal cost equals price. Ask yourself why this would not be maximising profits. Do not be disheartened if this proves too difficult at this stage. The answer is explained in the section Efficiency of Allocation.

2 That is the difference between what he earns per unit and his cost per unit.

The basic points are these:

- (a) In both a perfectly competitive and a monopolistic market a firm will maximise profits if it produces that level of output at which marginal cost equals marginal revenue.
- (b) In a perfectly competitive market the marginal revenue will equal price which the individual firm takes as given. So the firm will continue to produce until marginal cost equals price and the firm only needs to know its marginal cost to find its profit-maximising output.
- (c) In a monopolistic market the firm's demand curve is downward sloping, so that marginal revenue is below price. This means that when marginal cost equals marginal revenue it is also below price. The monopolist does not automatically know what output to produce by knowing his marginal cost curve. He has to know the shape of his demand and marginal revenue curves as well. In other words, the elasticity of demand for his product is a vital element in deciding his profit-maximising price and output.

3.3 Monopolistic Competition

A monopoly situation does not only arise when there is just one firm in a particular market. As was stated earlier some degree of monopoly exists whenever the goods of particular firms are not complete substitutes. So there might be a large number of firms in an industry, with each firm making its own product seem rather special in the eyes of the customer. This situation is known as product differentiation because consumers think there are differences between each firm's product. So a firm can raise its price and still be sure that it will not lose all its customers.

The market for many household goods is of this type. We all have our favourite toothpaste, for example, and even animals are not immune with many brands of cat and dog food each with their own supporters. This situation is partly competitive (there are lots of firms, and new firms can enter the market) and partly monopolistic (each product has a number of people wishing to buy it even if its price is higher). Not surprisingly the name for this is monopolistic competition.¹

3.4 Profit-Maximisation with Monopolistic Competition

In what way is the profit-maximisation behaviour of firms different in a monopolistically competitive market from a monopolistic market? The basic picture as shown in Figure 12.13 still stands. Firms still have some sort of monopoly; therefore they have a sloping demand curve; marginal revenue will therefore be less than price; marginal cost where it meets marginal revenue will be less than price and excess profits will be made.

But this is only half the story. A crucial difference between monopoly and monopolistic competition is that in the latter new firms can enter the market. It is relatively easy, for example, to

¹ After the work of Edward Chamberlin, who published his book *The Theory of Monopolistic Competition* in the U.S.A. in 1933. In the same year an English economist, Joan Robinson, published a very similar though independent book, *The Economics of Imperfect Competition*. Chamberlin's term seems to be the one most commonly adopted.

introduce a new brand of toothpaste. If excess profits are being earned by existing firms in the industry new firms will be encouraged to set up business. The result will be that they will take some business away from existing firms. For the individual firm its demand curve will shift to the left. It will be able to sell less output at any given price than it could do previously. This is shown in Figure 12.14.

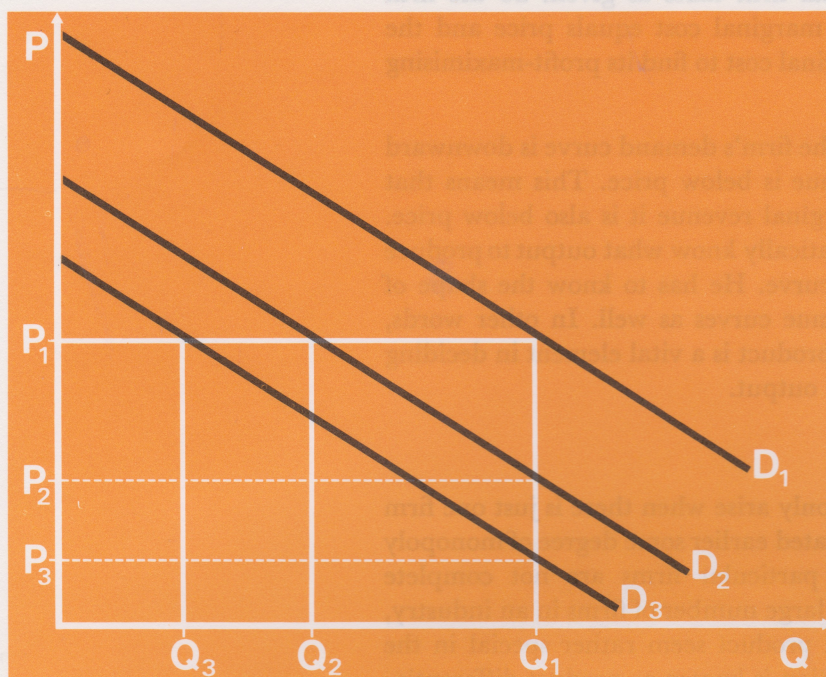


Figure 12.14

In Figure 12.14, D_1 is the original demand curve and we can assume that the firm earned excess profits. (The necessary MR, MC and AC curves to show this have not been drawn as this would make the diagram too cumbersome.) With this demand curve the firm could sell Q_1 at a price P_1 . New firms are encouraged to enter the industry. Existing firms are probably meeting the demand for the product so the result of a new brand will be to take some demand away from existing firms rather than increase total demand. The demand curve for the product of the firm shown in Figure 12.14 will shift to D_2 . If it sells at the same price P_1 it will only be able to sell the amount Q_3 . If it wishes to sell the same quantity as before it will have to reduce its price to P_2 to gain extra customers from its competitors. There may still be excess profits in the industry, in which case more firms will enter the industry and the demand curve of the firm will shift further to the left to D_3 . If it keeps its price at P_1 it will sell only the amount Q_3 . If it wishes to sell the amount Q_1 it must reduce its price to P_3 .

In the long run, in a monopolistically competitive industry, as long as new firms can enter the industry they are likely to reduce the demand curves for the products of other firms until no excess profits are being made. However, a certain firm's products may be regarded as so special by customers that new firms will not be able to reduce its demand and it will continue to make excess profits. This is not likely to be the general case however. In a monopolistic industry it was assumed that new firms could not enter the industry and so any excess profits continued to accrue to the monopolist over a long period. In a monopolistically competitive industry, as long as new firms can enter the industry they will usually eliminate any excess profits in the long run. The long-run profit-maximising

output will still be where marginal revenue equals marginal cost but at this point average revenue will also equal average cost. This is shown in Figure 12.15 below.

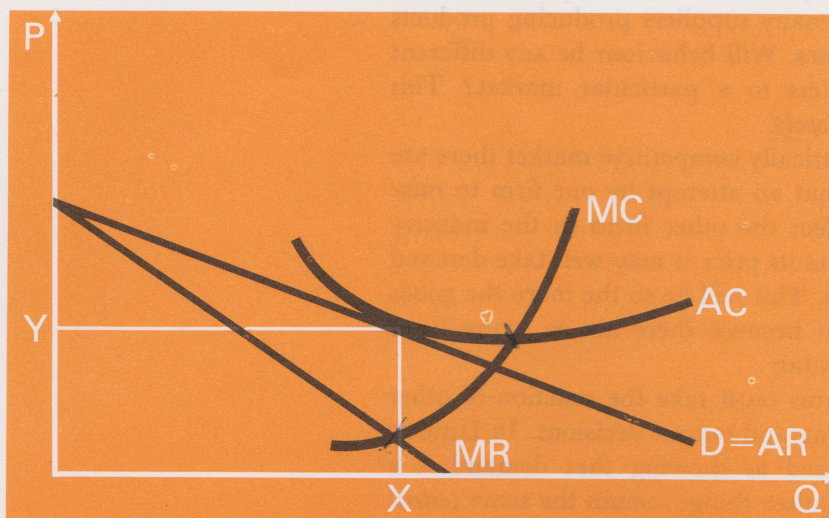


Figure 12.15

The demand curve $D=AR$ is similar to that in Figure 12.13. As explained in Figure 12.14 it has been pushed further and further to the left until it just touches the average cost curve. At this point the marginal revenue curve appropriate to the new demand curve crosses the marginal cost curve and output X , at price Y , is the long-run profit-maximising output and price. No excess profits would be made as average revenue (AR) equals average cost (AC) at this point.

The reason why this is a long-run equilibrium should be clear. If the demand curve were shifted below AC, in the long run the firm would make losses and go out of business and the demand curves of the other firms would rise as they took over its output until they again reached the AC curve. If the demand curve was above AC excess profits would still occur and new firms would enter the industry to reduce the demand curve until it reached the position shown in Figure 12.15.

This is essentially a *long-run* equilibrium position of a firm attempting to maximise profits in a monopolistically competitive industry. Only detailed observation of particular industries over a long time-period will show whether the equilibrium position is being achieved. It depends crucially on the ability of new firms to enter the industry and this explains why a great deal of energy and resources are spent by existing firms in an industry to keep new firms out. A large section of industrial economics is concerned with the barriers to entry of new firms in various industries. These barriers are of course not physical. Sometimes they may be legal or technical but usually they are financial. If firms have managed to make their products different in the eyes of the consumer by extensive advertising, a new firm will also have to spend a great deal of money to make consumers aware of its particular product.

The theory of monopolistic competition enables some of the complexities of industrial life to be incorporated into a theory of price and output behaviour, and helps to explain why firms in these types of industries behave the way they do. More particularly it helps to explain the role of advertising in these industries and the importance of new firms being able to enter them.

3.5 Oligopoly

The effect of the sloping demand curve on profit-maximising behaviour has so far been analysed in a market with only one supplier and in a market with many suppliers producing products which are not complete substitutes. Will behaviour be any different if there are only a few suppliers to a particular market? This situation is known as one of *oligopoly*.

In a competitive or monopolistically competitive market there are so many other manufacturers that an attempt by one firm to raise its share of the market will affect the other firms in the industry only marginally. If a firm reduces its price it may well take demand away from other firms' products. This will be so the more the goods are regarded as substitutes. But because there are so many firms no one firm is likely to suffer unduly.

In an oligopolistic market firms must take the reaction of other firms into account in their pricing and output decisions. In Unit 10 the demand curve was explained as showing that demand is a function of price, assuming all other things remain the same (*ceteris paribus*). Among the other things (*parameters*) that stayed the same were the prices of all other goods including those of a firm's competitors. With oligopoly a firm cannot assume that when it raises or lowers its price demand will increase or decrease. There are so few firms that if one attempts to increase the demand for its product this will have a marked effect on the demand for the other firms' products. The latter are likely to seek to preserve their share of the market. Other firms may therefore raise or lower their prices as well so that the price of one firm's product relative to that of its competitors is unchanged.

The effect of this is that the shape and slope of a firm's demand curve in an oligopolistic market cannot be stated with any certainty or precision. It cannot be sure how much output it will gain or lose by lowering or raising its price. A firm will still attempt to maximise its profits by producing at the point where marginal cost equals marginal revenue. But without knowledge of its demand curve it will not know its marginal and average revenue. In fact there is no one theory of oligopoly. What economists have done is to postulate different assumptions about how firms will react to their competitors' actions and then produced different theories for each different set of assumptions. Some of these come under the 'Theory of Games' which postulates certain strategies by a competing small number of parties.¹ This analysis can also be applied to problems such as International Relations where there is also a relatively small number of parties involved, with each not sure what the others' reactions to its own actions will be.

These theories can become extremely complex, but one point is clear. The basic problem in an oligopolistic situation is one of *uncertainty*. In an oligopolistic market therefore prices are likely to be stable. Firms cannot gauge the reactions of their competitors to any price changes they make, so it is usually in the interest of all firms to keep prices stable. They will prefer to compete in other ways – product differentiation through advertising, free gifts, competitions, etc.

¹ Theories of oligopoly will be more extensively discussed in second and third level courses.

The uncertainty of the effects of changing prices, together with the small number of other producers, means that tacit or open agreement between the different producers to keep prices at a particular level is often found in oligopolistic markets.

TEST 3

4 THE FUNCTIONS OF THE PRICE SYSTEM

Society can establish a number of different economic systems to decide which goods will be produced and how they will be distributed, and two alternative systems were described in Unit 3. Western society generally uses a pricing system¹ in which it is the prices which consumers are willing to pay that decides which goods will be produced. This unit finishes by considering how the analysis of equilibrium price and output behaviour throws light on the functions of the price system. Lack of time permits only a summary analysis but as supporting reading the student should study the extract by Phelps Brown & Wiseman in the *Reader* (pp. 215–23) entitled 'The Pricing System and Resource Allocation'. The subsequent paper (pp. 223–28) by Nove, 'The Soviet Economy: Micro-Economic Problems', provides a description of some of the problems of a different kind of economic system.

4.1 Allocation of Resources

One of the main functions of an economic system is to allocate the scarce resources of the economy to the production of goods. If a pricing system is adopted it must carry out this function. The pricing system achieves this by signalling to the resources of the economy (land, labour and capital) where they might most profitably be employed to produce the goods which members of the society desire. An organisation (called the firm) is established to hire and organise these resources to produce the goods. If consumers want more of a good than is being supplied the price system acts as a signal to the producers. The demand curve shifts to the right, price rises and suppliers allocate more resources to produce the goods. So price acts on supply and demand to match the amount consumers wish to buy with the amount producers wish to sell. Moreover price does not only act as a signal that supply should be raised (or lowered). It also specifies by how much it should be raised (or lowered).

Even in western economies not all resources are allocated by price. Some goods or services are known as 'public goods' in which price is not the deciding factor on how much is spent. The allocation of resources to external and internal defence is a political decision paid for by general taxation. Even here, however, some people pay for private security guards. A recurrent political issue concerns the area over which political decision rather than the price system should determine which goods are produced. Resources for education, for example, are mainly decided by the Government and paid for out of taxation. However, a private education system exists side by side with the public one in which the amount of resources devoted to private education is decided by the price system. The health

¹ Also known as a market economy.

service is another controversial area of political versus price-system allocation of resources.

4.2 Efficiency of Allocation

It is not sufficient to allocate resources. If the most benefit is to be gained resources must be allocated efficiently. This is usually taken to mean that resources are allocated to produce what consumers want as shown by the relative prices they are willing to pay. If resources have been efficiently allocated then they should not be able to earn more by switching to produce something else. This arises from the definition of costs given in Unit 11. The cost of using a factor of production is what that factor might have earned if it was used to produce something else. As long as factors could earn (and produce) more by doing something else they have not been efficiently allocated.¹

Resources will have been allocated most efficiently when they are producing the level of output at which the price consumers are willing to pay equals the marginal (extra) cost of producing it. If resources are to be allocated most efficiently in the economy as a whole then for every product the level of output must be that level at which marginal cost equals price. If for one product resources were allocated to produce a level of output at which marginal cost was less than price it would pay to switch resources to producing that good. If the level of output was such that marginal cost was higher than price this would indicate that the cost of producing the extra output was greater than the price consumers were willing to pay. Resources should therefore be switched from producing that product until output came down to the level at which marginal cost equals price.

But marginal cost equals price only in competitive market structures. The analysis in this unit has shown that the firm will maximise profits by producing to the point at which marginal cost equals marginal revenue. And at this point in competitive markets marginal cost equals price. The essential difference between allocation of resources in a competitive and a monopolistic market is shown in Figure 12.16.

In a perfectly competitive industry the demand curve $D = AR$ will be the industry demand curve. The marginal cost curve is the summation of all the individual firms' marginal cost curves and, as explained in Unit 11,² it becomes the supply curve for the industry as a whole. It cuts the industry demand curve at X at which the amount q_2 will be bought and sold at price P_2 . The price P_2 will be dictated by the market to all the individual producers who will attempt to produce as much as they profitably can at that price according to their individual marginal cost curves.

The same diagram can show the position for a monopolist. The demand curve $D = AR$ now becomes the demand curve for one firm as well as for the industry. The monopolist will also produce where marginal cost equals marginal revenue³ and this is output q_1 . At

¹ The assumption in this argument is that people's earnings reflect their contribution to output. So if people earn more they must have produced more. This assumption will be discussed in greater detail in Unit 16.

² See Unit 11, pp. 50–2.

³ See pp. 73–4.

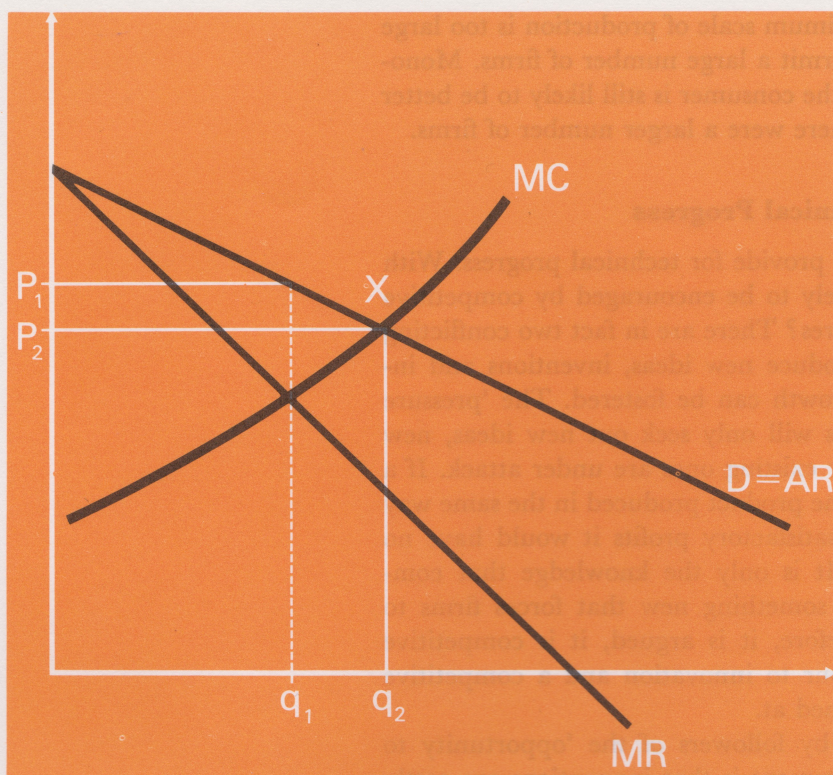


Figure 12.16

that output offered for sale the price will be P_1 . So with a monopolist the public is offered less for sale and at a higher price than with a perfectly competitive industry. This arises from the fact that in the perfectly competitive industry marginal revenue is equal to price for the individual firm so production is carried on until marginal cost equals price. In a monopolistic industry marginal revenue is less than price so that production stops before the point at which marginal cost equals price. The monopolist finds that he can maximise his profits at a level of output which is less than consumers wish to have. This is the essence of the classic economic argument against monopolies. Not that they are inherently inefficient or obsessed with power but that they result in the customer being offered less to buy and at a higher price. For the economy as a whole there is an inefficient allocation of resources to the production of that particular good.

It would seem therefore that resources will only be allocated efficiently in a perfectly competitive market. There are, however, very few perfectly competitive industries in the United Kingdom. Once a departure is made from perfectly competitive assumptions it does not automatically follow that a market structure slightly more competitive will produce a more efficient allocation of resources than one slightly more monopolistic.

The assumption in Figure 12.16 is that whether an industry is perfectly competitive or monopolistic it will still have the same cost curves. This of course may not be so. In certain industries substantial *economies of scale* exist so that average costs per unit of output decline as output increases. It may well be cheaper to have the total industry output produced by only just a few firms with large production plants than by a large number of small firms with much smaller production plants. A study of the oil refining industry has shown that when the capacity of a refinery is increased ten times the operating costs can fall by over 50 per cent (Pratten and Dean, 1965, p. 91). In cases such as this, less than perfect competition is

bound to exist because the optimum scale of production is too large in relation to the market to permit a large number of firms. Monopoly profits may be made but the consumer is still likely to be better off in terms of prices than if there were a larger number of firms.

4.3 The Price System and Technical Progress

An economic system must also provide for technical progress. With a price system is this more likely to be encouraged by competitive or monopolistic market structures? There are in fact two conflicting viewpoints on how best to produce new ideas, inventions and innovations so that economic growth can be fostered. The 'pressure to innovate' view is that firms will only seek out new ideas, new products and processes, if their existing ones are under attack. If a firm could go on selling the same product produced in the same way year in, year out, and make satisfactory profits it would have no incentive to seek new ideas. It is only the knowledge that competitors may be bringing out something new that forces firms to think about the future. Therefore, it is argued, it is competitive pressure which is the main spur to innovation and a competitive market structure should be aimed at.

This argument is attacked by followers of the 'opportunity to innovate' viewpoint. Their argument is that innovations are costly to introduce and involve such a large amount of resources being devoted to research and development that only large firms can afford to finance them, usually from monopoly profits. In addition, Joseph Schumpeter (1947) argued that it is only the prospect of large profits that persuades firms to undertake the large risks involved in producing a new product or process. In the long run, it is argued, society benefits from large firms having some form of monopoly over the market, for only in this way will there be progress. With competitive industry there may be short-run benefits in cheaper and more plentiful products but eventually there is stagnation. Monopoly profits become almost like a tax on consumption which is devoted to investment in the future. The drug companies in this country often argue that they need to make large profits out of the National Health Service in order to devote resources to research and development to produce new drugs.

This is not the place to discuss this issue more fully but it serves to show that a case can be made out for monopoly profits. In fact current views of the innovation process combine both arguments. It is now argued that it is *international* competition which produces the pressure to innovate. If we wish to compete in world markets we must have the best products and most up-to-date processes, otherwise we shall lose business. To do this, it is argued, we must have large companies capable of matching the U.S.A. giant corporations and able to devote large sums to research and development. The arguments in favour of the G.E.C. merger with A.E.I. and English Electric in 1968 were of this nature. If there is any lesson to be learnt from all this it is that every case must be judged on its own particular merits. There is no automatic rule about competition and monopoly.

4.4 Limitations of the Price System

The fact that the price system may only ensure an efficient allocation of resources in a perfectly competitive market has already been

mentioned. Its ability to ensure that resources are so allocated to maximise the satisfaction of people's wants is open to a number of other criticisms. These were first mentioned in Unit 3¹ and the relevant section should be read again. The main points are summarised below with some additional arguments in a number of cases.

I RATIONING AND THE DISTRIBUTION OF INCOMES

The price system assumes that people signify their wants by the amount they are willing to pay. The more they want a good the more they are willing to pay. So if there is a shortage of a particular good the price will rise until the amount demanded equals the amount supplied. At this price the goods will go to people who want them most, as signified by the amount they are willing to pay.

This method of maximising people's satisfaction may seem fair at first sight. However, the price people are willing to pay for a good depends to a large extent on their incomes and wealth.² If it is 'fair' for price to decide which people get certain products then the manner by which people obtained the means to obtain these products must also in some sense be 'fair'. In other words if the price system is accepted as the correct method by which goods are allocated to consumers this implicitly assumes that the distribution of national income between different sections of the community is also correct. Unit 16 will discuss the distribution of income but it may be briefly stated that one of the ways in which distribution is thought to be correct is if earnings reflect the person's contribution to output. This is only thought likely to exist if there is perfect competition in the market for labour. However, with monopoly employers and monopoly unions, distortions are bound to exist. Questions of what is 'best' and 'fair' come within the sphere of a more advanced area of economic study called 'welfare economics'.

The price system, of course, is not the only method of rationing. In time of war when consumer goods are particularly scarce the Government does not allow the price system to carry out its rationing function. Instead each person is allowed to receive a specified amount by means of a points system. This usually produces an unofficial price system in the form of a 'black market'. Whenever the price system is prevented from carrying out its rationing function in an official market it usually re-appears in some unofficial market.

II UNEMPLOYMENT

The price system only produces an efficient allocation of resources if, in fact, it ensures that all resources are used. If there is unemployment then, of course, all resources cannot be efficiently utilised, as some are lying idle. The unemployment of the 1930s showed that the price system does not automatically ensure a full employment of all resources. The theories of Lord Keynes showed that Government intervention in the market economy was necessary to ensure full employment. These theories will be discussed in Unit 14.

1 Pp. 74-78.

2 See Unit 10, pp. 20-22.

III PRIVATE AND SOCIAL COSTS

If the price system is to produce the most efficient use of resources from the point of view of society as a whole then prices should reflect the social costs of producing the good and not just the private costs that the firm alone has to bear. The classic example of this is the factory which in the course of manufacturing its products produces smoke which dirties the clothing drying in the gardens of nearby houses. The extra washing of the clothes is clearly a cost of manufacturing the product. But it is not a cost that the factory has to pay. If the factory had to pay the full social costs of manufacturing the product these costs would be higher, social marginal cost would equal price at a lower output and less would be produced. So if the price system is to ensure that resources are efficiently allocated these resources should be employed on the basis of their social marginal cost and not their private marginal cost.

IV RIGIDITIES IN THE SYSTEM

For the price system to work effectively resources must be flexible in adjusting to consumers' demands. If consumers want less of a good, its price falls, marginal cost exceeds price, and resources are released and transferred elsewhere. In reality rigidities exist. A firm sometimes does not sack workers if demand falls off. Capital is even more inflexible. A piece of specialised equipment cannot be switched to producing something else if the demand for the original product is reduced.

These sort of rigidities in capital and labour mobility are sometimes used to justify government support for declining industries. The fact that it is difficult to find an alternative use for specialised capital or for middle-aged workers with specialised skills means that these resources might be most efficiently used in being subsidised to produce what they are best qualified to produce. The alternative may be unemployment, in which case no output at all is produced.

V MANIPULATION OF WANTS

The assumption that the price system allocates resources to maximise the satisfaction of people's wants implies that the allocation of resources itself cannot change people's wants. However, with advertising, people can be made to want what resources have been allocated to produce and not vice versa. The question of whether 'demand creation', as it is known, leads to a more efficient allocation of resources, is a very complex one. At the very least, however, it throws doubt on the function of the price system.

So there are a number of limitations to the ability of the price system to ensure that resources are efficiently allocated. This is not to say that an alternative system could produce a more efficient allocation. Such alternative systems as a command economy¹ and a mixture of resource allocation by price and command systems need to be examined and analysed.

¹ For a definition and description of a command economy see Unit 3, pp. 79–80.

5 FACTORS OF PRODUCTION AND THE PRICE SYSTEM

The analysis has concentrated on the way in which the price of *goods* is determined, but the price system is also used to equate the supply and demand for factors of production. The price of labour is shown by the wage rate. In a perfectly competitive labour market the wage rate would equate the supply of labour with the demand. However, with monopolist hirers of labour (i.e. firms) or monopolist sellers of labour (i.e. unions) this may not be the case. The price of capital is given by the rate of interest. This reflects the willingness of owners of capital to lend and the willingness of hirers of capital to borrow. If more people wish to borrow than wish to lend the rate of interest rises and if more people wish to lend than to borrow then the rate of interest falls. The factors that affect people's willingness to lend and borrow capital will be explained more fully in Units 14 and 15.

TEST 4

6 CONCLUSION

The analysis presented in the last three units has assumed that individuals will try to achieve maximum satisfaction from their purchases, and that firms will attempt to maximise profits. But economic organisations are also social organisations and economists are well aware that in real life economic behaviour is also affected by social forces which are not themselves primarily economic. To underline the role of these social forces the next unit will discuss the sociology of economic behaviour.

BIBLIOGRAPHY

- CHAMBERLIN, E. H. (1933). *The Theory of Monopolistic Competition*. Cambridge, Massachusetts, Harvard University Press.
- PRATTEN, C., and DEAN, R. M. (1965). *The Economics of Large-Scale Production in British Industry*. Cambridge, Cambridge University Press.
- ROBINSON, J. (1933). *The Economics of Imperfect Competition*. London, Macmillan.
- SCHUMPETER, J. (1947). *Capitalism, Socialism and Democracy*. London, Allen and Unwin.

Unit 13

The Sociology of Economic Behaviour



CONTENTS UNIT 13

	PAGE
1 INTRODUCTION	91
2 THE FAMILY AND THE ECONOMY	91
2.1 Industrialisation and the Family	91
2.2 The Woman Worker	93
2.3 The Persistence of the Kin Unit in Economic Organisation	94
3 SOCIAL VALUES, CULTURE AND ATTITUDES	94
3.1 The Definition of 'Social Value' and 'Culture'	95
3.2 Social Values and Economic Performance	95
4 THE SOCIAL FOUNDATIONS OF PROPERTY	97
4.1 The Nature of Property Rights	97
5 EXCHANGE AND CONSUMER BEHAVIOUR	98
5.1 'Gift Exchange'	99
5.2 Contract	99
5.3 Reciprocity, Exchange and Power	100
5.4 Consumer Behaviour	100
6 LABOUR: WORK AND OCCUPATIONS	101
6.1 The Concept of the 'Net Advantages'	101
6.2 Leisure and Work	103
6.3 The Drive for 'Autonomy' in Work	103
6.4 The Goals of Occupations	104
7 CAPITAL AND ENTERPRISE: THE GOALS, OWNERSHIP AND CONTROL OF COMMERCIAL ORGANISATIONS	105
7.1 Business Goals	105
7.2 Ownership and Control	105
8 CONCLUSION: ACHIEVEMENT AND ASCRIPTION IN THE ECONOMY	107
BIBLIOGRAPHY	108

THE SOCIOLOGY OF ECONOMIC BEHAVIOUR

I INTRODUCTION

In recent weeks you have been introduced to some of the basic ideas of economics. Earlier, in Units 5 and 9, you were presented with some basic ideas in sociology. In this unit we are going to consider some ways in which these two approaches fit together.

When economists are talking about one area of the economy – such as the price of certain commodities – they may deliberately exclude from the argument some possible economic changes elsewhere which could affect the prices. ‘Other things being equal’ is a phrase by which not only economic factors can be deliberately excluded – such as the price of some other commodity; ‘other things’ which affect prices might also be social factors – such as a change in fashion.

Economic co-operation, or the division of labour, is obviously not purely an economic matter; it also involves social co-operation. The great French sociologist, Emile Durkheim, pointed out in his book *The Division of Labour* (1893) that the development of more specialised forms of industrial production go hand in hand with very profound social changes – such as the mushroom growth of cities and changes in family life.

Sociologists are therefore interested in various aspects of society in which the economic element is obvious and prominent. Industrial sociologists, for instance, study what happens in factories and other sorts of industrial situations. Other sociologists study occupations, professions, and trade unions. Social class and social stratification – another area where income and other economic differences are obviously important – are prominent topics in sociology,¹ which will be discussed in Unit 17.

I want to begin by mentioning how one individual sociologist, Talcott Parsons, sees society as made up of ‘sub-systems’. One of these sub-systems of society is the economy, which he calls the ‘adaptive’ sub-system. Another sub-system is that which handles the functions of *pattern maintenance* and tension management – most of this is done by the human family. In the next section we see how Parsons analyses the connections between these two sub-systems – between the economy and the family.

2 THE FAMILY AND THE ECONOMY

2.1 Industrialisation and the Family

Every economic system has connections with non-economic aspects (or ‘sub-systems’) of the same overall society. In non-literate societies, before the arrival of large-scale manufacturing and other industry, the economy is deeply embedded in the rest of the social

¹ Sociology as a (social science) discipline is concerned with the whole of society; there is not only a sociology of economic behaviour but also of political behaviour, religious behaviour and so on. Such broad concerns lead some people to argue that sociology is very broad and also very shallow. However, perhaps for the moment we can leave aside the question of whether sociology is the top dog or the stray mongrel of the social sciences – as well as the third possibility that it is neither.

system. It is often difficult to tell where the economy ends and the family, for instance, begins. This is because economic activity typically takes place within a family context – on family farms and in small home-based industries. Similarly the economy and the political system are very much ‘all part of the same thing’. Such societies are called ‘primitive’, ‘non-literate’, or ‘simple’.

When a society industrialises – adopts a complex technology, a central government, and modern education – then the family system tends to play a smaller (but still often a substantial part) in political and religious, as well as economic life. In such an industrialising society specialisation and the division of labour is given increased emphasis. Economic units – such as mines and factories – develop quite separately from the context of family and kin.

An intermediate type of economic unit common in the early stages of industrialisation is the family-based economic unit supplying goods or services to a distant market. Examples of such kin-based economic specialisation include the peasant non-subsistence farmer who grows a cash crop; simple craft industries in which the family stays at home to produce the goods (the ‘domestic’ system) which are nevertheless marketed outside the local area. Yet a third example is the family-supported trader or market-stall operator.

In a more ‘advanced’ form of industrialisation involving a more extensive division of labour, the work sphere becomes completely removed from the family sphere. The peasant farmer’s son may go to work in a distant mine or factory, often leaving wife and children behind. Work and family then become geographically separate. When this first happens the social implications are profound and the dislocation in the lives of individual family members often immense. Individuals find difficulty in adjusting to long periods of months or even years spent working at and living near the factory or mine, interspersed with brief periods in the rural home of their parents. At ‘home’ social relationships are based on close kinship, usually life-long familiarity, and deeply felt emotions (love and perhaps also hate). In contrast social relationships at ‘work’ are based on ‘rational’ and ‘financial’ criteria in which inter-personal emotions, while obviously not absent, are much less basic.

This is clearly an extremely simplified account of what happens to the family and economic spheres with industrialisation. Nevertheless, as Talcott Parsons has pointed out, the connections between family and work in an industrial society involve contrasting values which have important implications for the society. In Parsonian terminology the emotional contrast between the family and work involves a contrast between *affectivity* (family affections and other emotions) and *affective-neutrality* in the case of work relationships based largely on economic rewards. Another contrast is that family relationships are *particularistic* (confined to particular individuals such as one’s own children) – whereas at work *universalism* prevails; that is to say, individuals are judged upon such universal or ‘rational’ criteria as educational achievement, measured skills and other ‘impersonal’ attributes.

Obviously these opposed spheres of family and economy are not, and cannot be, utterly separate. Mr. Jones at home with his wife and children on Sunday and Mr. Jones with his fellow workers at the factory during the week is still the same Mr. Jones. Mr. Jones takes his family loyalties with him to work – and Mr. Jones also takes home from work not only his pay-packet but also a variety of

social values, anxieties and hopes which derive from work.

Between the contrasted spheres of family and work, various choices confront the individual, the family, the employer, and society at large. The choices which confront the individual are embraced by the economist's concept of the 'net advantages' of different kinds of work. 'Net advantages' might include the remuneration in both money and non-money (e.g. meals) forms; the pleasantness or unpleasantness of the work; and such factors as social interaction with work mates. 'Net advantages' are strongly influenced by family considerations, for any individual at work must make choices involving his family. Should he choose a pleasant but badly paid job even though this means less money for his family? Should he work far from home if this means that he has less time and energy for his family life? Should he accept a job which his family disapproves of? Should he work overtime – which will mean more money for his family but less time with them (of course if he wants to get away from his family, overtime is obviously advantageous!)? Should he go on strike when he has children to feed?

2.2 The Woman Worker

Moreover, many employed workers are women.¹ The whole phenomenon of women and paid work cannot be approached at all without reference to the women's family circumstances. The most obvious point here is that a much higher proportion of single than married women go out to work; and there is a low proportion of full-time workers among the mothers of young children. The whole question of women going out to work, is, of course, a controversial one: a working woman with a husband and children has two major roles to perform – that of paid worker and that of wife/mother. Those who disapprove of married women going out to work tend to argue that these two roles are in conflict. Those who argue the other way often say that the roles can be complementary – that a woman may be a better wife and mother if she also goes out to work. Empirical evidence about working women suggests that many housewives see 'going out to work' primarily as an opportunity to earn extra money which can be spent on the family. But the evidence includes other complexities; the impact on the family depends upon the attitudes of husband and wife, and also upon a number of other variables. The phenomenon of the working wife is sometimes associated with marital dissatisfaction; however, some research confirms that the wife who wants to work, but does not, is also dissatisfied.

Whether or not a woman goes out to work usually involves a collective choice by several members of a family. If the wife/mother is to go out to work, several adjustments may be involved – by the husband/father; by children of various ages, and perhaps by grandparents in the form of baby-sitting. The choice also depends on the availability of various services in the local area (transport, shops, nursery schools). More basic still – women obviously cannot go out to work unless there is work for them to do. Some areas have very little work for women – while in others a high proportion of women

¹ Economists, incidentally, do regard housewives as productive workers – although their production lies outside the market economy: (the problem is to set a monetary value on housewives' work).

have traditionally been involved in paid employment. Among other things this suggests that the economic role of the woman with a family is extremely flexible.

The question of working women involves choices not only for the individual woman and for her family but also for commercial companies and other employing organisations. Some companies make special provision for women workers – including a flexible policy designed to enable women to stop work temporarily for pregnancies and family crises. Other companies prefer to employ women only in clerical or other ‘staff’ capacities.

Employing organisations have differing attitudes not only towards employing women but also towards the family life of men employees. Some organisations deliberately try to control the contrasted values and loyalties of family and work by encouraging wives to support the husband’s working role. An international company which depends upon being able to move its employees between different countries has a special interest in discouraging wives from becoming geographically fixed in permanent employment, since this will reduce the mobility of their husbands. Some such large organisations have a deliberate policy of not employing both a man and his wife; such a policy is intended to stress that a large organisation must emphasise ‘merit’ (universalism) and must discourage possible ‘nepotism’ (particularism).

2.3 The Persistence of the Kin Unit in Economic Organisation

In some types of economic organisation the kin unit remains, even in industrialised societies, the most common basis. This is true of many types of small business (shops, restaurants, small-scale services and farms). In the small business two major factors which favour the kin-based organisation are firstly the need for flexibility and motivation and secondly the pattern of inheritance. Economic activities such as farms, small shops and restaurants involve very long hours, ‘split shifts’, an uneven seasonal flow of work and so on. The immense flexibility and strong motivation of the wider kin group (especially the part-time support of women and children) is important here. Another aspect is the difficulty of raising enough capital. These problems are often solved, for instance, by combination between brothers.

Not surprisingly, then, much legislation is concerned with the place of families in the economy. The laws relating to the system of inheritance itself are one example. Any political power which wishes to alter the balance of economic power may be expected to make changes in the area of inheritance. For instance, feudal lords may be interested in maintaining a system of primogeniture whereas a king may try to break up large landholdings by requiring equal distribution of inheritance between all of a lord’s children. Political parties which want to redistribute wealth in society often choose to emphasise ‘death duties’ and similar taxes on the inheritance of property.

3 SOCIAL VALUES, CULTURE AND ATTITUDES

The object of this section is to consider the point that different societies emphasise different *social values*. For instance different societies may have different attitudes towards economic growth. By way of example we shall, in a moment, consider whether Britain

emphasises different social values from those emphasised in Japan and the United States – and, also, whether this may be related to different levels of economic performance between these societies.

3.1 The Definition of 'Social Value' and 'Culture'

First, however, the term *value* itself requires some brief explanation. How does 'value' as used by an economist differ from the same term 'value' used by a sociologist? 'Economic' values are involved when goods or services are scarce relative to the desire for them, or when their production requires the use of scarce resources. Economic values depend, therefore, upon the operation of supply and demand; and relate to goods or services which are in demand. When goods or services are marketed, then their value can be translated into price.¹ The sociologist, however, uses the term value more or less as it is used in everyday speech – and as it is used by philosophers or political historians. The sociologist talks not only of economic or utilitarian values – but also, for instance, of religious values, political values.

Within sociology the terms 'value' and 'norm' are sometimes used interchangeably. Both 'values' and 'norms' refer to what is regarded with approval (or disapproval), as desirable (or undesirable) or as meritorious. But the term 'norm' is used more frequently at the micro or small-scale level – for instance, norms of collegially behaviour among co-workers. The term 'value' is used more frequently at the macro or large-scale level – for instance, to describe 'democratic' values prevalent within a whole society.

Culture, as defined by sociologists, consists of many different sorts of values – not only artistic or aesthetic values and economic or utilitarian values, but also religious, political and perhaps especially what are often called 'moral values'.

3.2 Social Values and Economic Performance

The contention that social values and the broad social culture of a nation influence its economic performance is one which economists themselves often put forward. J. K. Galbraith has merely put this point of view with special pungency in his dictum: 'We are profoundly conditioned by the theology of the market.'

Do social values prevalent in Britain explain the (allegedly) weaker performance of the British economy compared with that of the United States, during the last 100 years? An American sociologist (Burrage, 1969, p. 117) has examined some of the evidence:

Explanations of Britain's low rate of economic growth are frequently sociological rather than economic. They refer either to British attitudes (amateurism, conservatism) or to particular institutions (trade unions, public schools, civil service) or perhaps to the British class system rather than to specifically economic variables. . . . Levine compared British and American economic performance 1880–1914. . . . He concludes 'the technical and organisational lag in British industry was more than anything else a question of entrepreneur responses; responses that drew their character from certain social and social-psychological circumstances. . . .'

¹ See Unit 10.

The author of this article, Michael Burrage, went on to examine the evidence on six organisational differences.

- 1 The degree to which companies use systematic and impersonal methods of selection and emphasise measured performance rather than age in promotion decisions.
- 2 The prominence given to scientists.
- 3 The degree to which communication patterns within the companies were open and 'democratic' rather than 'closed' and hierarchical.
- 4 The precision with which work roles were defined.
- 5 The extent to which management consultants and available sources of information and advice were used by companies.
- 6 The extent of executive mobility.

Burrage argues that less emphasis is given to all six points in British industry as compared with American industry. He believes that these differences are relevant to the greater growth of the American economy. Burrage himself admits that some of the evidence is unsatisfactory. But the broad argument that social factors make a contribution to differential economic growth rates seems highly persuasive.

Another example of how social values and culture influence the economy is provided by Japan. Here we can refer to the piece in the *Reader* (pp. 229–35) – 'Labour-Management relations in Japan' by the Japanese social scientist Shizuo Matsushima. The very first paragraph of this piece makes clear a remarkable fact about Japanese trade unions – 90 per cent are 'enterprise' unions. National trade unions are, by contrast, weak. Perhaps the most startling aspect of industrial life in Japan – as seen by western eyes – is the great emphasis given to job security. This emphasis leads inevitably to a stress on the importance of *recruitment* – hardly surprising where an employee once employed has a job for life. This emphasis on recruitment, however, leads to two highly contrasted types of recruitment policy:

- 1 *Executive* personnel are recruited on the basis of competitive examinations (the stress here being upon 'rational' procedures, universalism, and positions being earned on the basis of achievement).
- 2 *Non-executive* employees are recruited on a quite different basis – here family ties with existing employees are one major criterion and a way of ensuring positive motivation (the stress here being upon particularism and ascription – i.e. the job is acquired partly on the basis of birth, not individual qualities or merit).

The sharp distinction made between recruiting executives and non-executives is another example of accommodation between family/kin and work/employment. But it also implies an extremely hierarchical and paternalistic view of industrial relations. The remainder of Matsushima's article makes clear that such values are indeed prevalent in Japanese industrial relations. Length of service with the company is given extraordinary emphasis. Welfare services – including health and housing – are provided. And, of course, the predominant pattern of 'enterprise unions' fits into the overall framework of what Matsushima calls 'managerial familism'.

Consequently, 'firms have come to regard their labour costs almost as a component of fixed costs'. Clearly such a system provides not only new flexibility – compared with western arrangements – but it also makes for new rigidities. One obvious solution to the problem of having a fixed-cost labour force is that rapid expansion

will help to solve many internal company problems which would otherwise arise.

Obviously I am not suggesting that the success of the Japanese economy since 1945 can be so easily explained. But I would contend that any fully adequate account of the Japanese economy's performance could not possibly ignore major differences in social values between Japan and 'comparable' western nations.

The extent to which leadership roles in the economy are valued is also important for the economy. Does the society accord high status to its businessmen, its managers and the entrepreneurial function generally? Or does the society value more highly the lawyer, the intellectual, the artist, the soldier or the religious leader?

To some extent the educational process in any society merely reflects the values of the society and hands these values on to the next generation. But in any highly industrialised and highly specialised society education is to some extent a social system in its own right. Of course, any educational system is supported and financed by the existing economic and political order. But this very circumstance – that education is usually subsidised and does not have to 'pay its way' in the economic market place – further ensures that the values prevalent within the education system will be at least partly opposed to the values of industry and commerce. In particular this type of expensive and specialised educational system depends heavily upon the specialised occupational role of teacher. There is a notable tendency, in many nations, for the teaching occupation (or profession, or sub-profession) to be somewhat opposed to the values of the business world. Teachers not infrequently see their own motives as purer and less self-seeking than those prevalent in industry and commerce. On the other hand, people who themselves work in industry and commerce may regard teachers as living in ivory towers, and being cut off from reality.

The importance of the educational system for shaping the overall values of a society has been known to the Roman Catholic church for many centuries. Communist regimes also give special emphasis to education. Nor has the significance of the educational system escaped the notice of business men in western societies during the twentieth century.

4 THE SOCIAL FOUNDATIONS OF PROPERTY

The concept of 'property' appears central to the economic order of any society. Those who wish to make fundamental criticisms of any political and social order usually challenge either the extent of private property – especially the ownership of capital by individuals – or the extent of public property, such as public ownership of particular economic services or industries.

However, one must not make the mistake of believing that such terms as 'property' or 'ownership' carry one precise meaning. In some societies a man may 'own' another man. In other societies the state 'owns' all major industrial and commercial enterprises. In many societies a man may 'own' a house. The concept of property or ownership refers to many different sorts of rights, exercised by different sorts of people and groups in relation to different goods, services and abstractions.

4.1 The Nature of Property Rights

Individuals within a society may have a variety of different sorts of

rights – some of which may be referred to as property, and some of which may not. One sociologist (Johnson, 1961, p. 225) lists seven kinds of possible right:

- | | |
|---------------|----------------|
| 1. Possession | 5. Usufruct |
| 2. Use | 6. Income |
| 3. Alteration | 7. Disposition |
| 4. Using Up | |

Rights to *possession* and rights to *use* are very similar – but need not include the right of alteration. The man who rents a house acquires for a period rights of ‘possession’ (or occupancy) and of use, but he does not usually acquire the right to knock down and re-build. The right to *alteration* is one which the house owner (still with some limitations) does acquire.

The right to *use up* is most obviously illustrated by ‘mineral rights’ – such as the right to mine coal or pump away oil reserves. *Usufruct* or the right to the produce of land is also somewhat different. A man who rents a farm may nevertheless sign a contract with a food processing company which hands over the right to harvest and remove all the vegetables from a particular area.

Another kind of right is the right of the author of a book to an *income* or royalties for the publication of that book. This kind of right – which is also acquired by musicians and artists of various kinds – involves property of a very abstract kind. The seventh kind of right is *disposition*. Here the owner acquires the right to destroy the property in question – for instance, by eating a piece of food which he owns, by burning some fuel he has bought, and so on.

The variety of different sorts of ‘property’ obviously follows partly from the very wide range of physical objects, and services which have an economic value. But the whole constellation of ‘property’ rights can vary radically from one society to another. We have already mentioned obvious differences between a market economy and a command economy. The contrast with medieval Europe is equally sharp. The relations between a medieval lord and his vassal are an obvious case. For instance, a vassal might ‘own’ property, but the property might nevertheless revert to the lord at the vassal’s death. Another medieval limitation was that the vassal might be allowed to hand on his property at death – but was restricted by a law of primogeniture.

The institution of ‘property’ is basic to the structure of any society – on this the policeman and the anarchist, as well as the capitalist and the communist, can probably all agree. The implication that the grant of rights to particular individuals in relation to particular objects must deny those rights to other individuals also seems unarguable.

‘Property’ defines the legal rights of people in relation to things which have a scarcity – or economic – value. When one man ‘owns’ a particular scarce thing, other men do not. Consequently, property relations are relations between people. The distribution of property within a society defines social and power relations between people and between groups of people.

5 EXCHANGE AND CONSUMER BEHAVIOUR

Economic specialisation within any society involves some division of labour, and division of labour cannot exist without some system of exchange. This system of exchange is required in order that

different individuals and groups may reap the benefits of their specialisation by exchanging their specialised goods and services in return for the different specialised goods and services of other producers.

Division of labour and exchange are then social as well as economic phenomena.¹ While such a crude exchange process as barter is obviously a social process, the fact that advanced economies use money as a central device in exchange does not eliminate the play of social forces. This is especially obvious in consumer behaviour. Shopping – whether conducted in a small local store, or a self-service supermarket, or in a large department store – is clearly a social activity. Clearly also such factors as prestige are of great importance.

5.1 'Gift Exchange'

In societies which do not have a money economy – or in areas of developing countries which lie largely outside the money economy – there are often traditional patterns of exchange. One pattern which has been of particular interest to social anthropologists is 'gift exchange'. Where this type of arrangement exists there is a strong social obligation for one individual to give goods or services to another individual who stands in a certain relationship to him; if the first individual fails to make this gift he loses prestige – but if he gives especially generously he may gain extra prestige. Such an arrangement may seem at first sight grossly illogical. But, the principle of *reciprocity* seems to be very firmly based in all human societies. When gifts are given they establish obligations for the gifts to be returned. And in societies where gift-giving is a prominent feature of social life gifts tend to be reciprocated. Sometimes the element of reciprocity operates directly, and in other situations it works indirectly.

In some simple societies the goods and services which are given and received are so similar that the economic basis of the exchange appears small. In these circumstances the major functions of gift-giving appear to involve firstly *friendship* – gifts cement bonds with other individuals. Secondly, such giving and exchanging has an element of politics and *power*; for instance, the man who gives especially large amounts establishes superiority. In other types of gift exchange the goods given and received may have a more clear element of economic exchange; this may involve the exchanging of different goods, or there may be an element of investment over a period of time in return for a dividend or interest on the investment when the gift is exchanged.

5.2 Contract

In an industrialised society, gift-giving, although less prominent, still exists, and the element of reciprocity still remains. In an industrialised society there is not only the device of money (and more elaborate markets than those found in peasant societies), there is also the institutionalisation of exchange in terms of 'contract'. Contract like 'property' is backed by the legal apparatus of the society in question. Like property, contract also carries with it important elements of potential conflict, of power relationships – and also all kinds of social overtones. Contracts, of course, cover

¹ See Unit 3. 'The division of labour depends on the extent of the market'. (Adam Smith.)

many relatively small transactions. But there are also contracts which cover very large transactions – for instance, between two large commercial organisations. Such a contract – which may relate to the building of a factory, or a jet airliner – is of great interest to sociologists – and especially to that branch of sociology called the sociology of law. One interesting aspect of these contracts is that although they have full legal validity, the parties very seldom attempt to enforce the contract in a court of law. Even when one company fails to build the ship or the factory on time relevant clauses in the contract may not be enforced. Resort to the law tends to be held in the background as a possible ultimate sanction – which is all the more powerful for being seldom invoked. But another aspect is that individual members of the two commercial organisations will weave between them a network of personal relationships – through which disputes can be adjusted, and various forms of reciprocal ‘fair play’ established and carried out.

5.3 Reciprocity, Exchange and Power

Sociologists like Peter Blau (1964) have argued that much of what occurs in social life can be seen in terms of reciprocity, exchange and power. Elements of exchange are present between colleagues working in the same organisation, between marriage partners, and in small groups of all kinds. This perspective of exchange may help to explain what sometimes appears to be inexplicably generous and spontaneous behaviour. Possible examples of this occur in the piece in the *Reader*, by D. Roy (pp. 235–45), called ‘Making-out: A counter-system of workers’ control of work situation and relationships’. Roy himself says (p. 244):

No attempt has been made here to account for the aid extended by service groups, though suggestion that this assistance might be part of a larger system of reciprocal obligations has been implicit. It is apparent, for instance, that tool-crib attendants benefited from their practice of admitting operators and set-up men to their tool room to seek and pick up equipment.

5.4 Consumer Behaviour

Consumer behaviour in some ways comes before production. The whole field of consumer behaviour is of increasingly great importance, not only for the economy but for the central part it plays in human life in general. Some of the possibilities of this area can be illustrated by the following quotation from a study about the shopping habits of middle class housewives in Chicago in the 1950s (Musgrave, 1969, p. 55):

Thirty-three per cent were called ‘economic’, since they showed a detached manner. They were young, aspired to high status and saw themselves to be at an economic disadvantage; they had few allegiances in the area in which they were living. There were, therefore, few social constraints on their behaving in a rational manner as consumers. Twenty-eight per cent were ‘personalising’, having close relationships with local storekeepers. They seemed to be newcomers of low status, seeking some roots in the area. These consumers were willing to exchange somewhat dearer goods for other rewards such as friendliness. Eighteen per cent were ‘ethical’ shoppers. These had lived in the district for a long time, were of higher status and felt they ‘ought’ to shop there despite the cost. Finally, apart from the few who could not be categorised, seventeen per cent were ‘apathetic’, aiming to eliminate effort in their shopping rather than to seek bargains. Very often these were old people.¹

¹ Musgrave is summarising another study: G. P. Stone, ‘City Shoppers and Urban Identification’, *American Journal of Sociology*, July 1954.

The general point which this quotation makes is that empirical research supports the everyday observation that people's consumer behaviour is strongly influenced by their social circumstances and social attitudes.

6 LABOUR: WORK AND OCCUPATIONS

The most obviously 'social' of all the economist's factors of production is labour. Needless to say, since it is rendered directly by people, the labour factor does not always respond to market forces in the way suggested by a purely economic model. On the other hand, some over-sophisticated or (excessively naive) social scientists have sometimes assumed that money is less important for employees than satisfactory 'human relations' at work. The 'human relations' school in sociology prospered in the 1940s but few sociologists now take it too seriously.

6.1 The Concept of the 'Net Advantages'

It is now usual to regard work as for most people a *primarily* money earning or 'instrumental' activity – in which certain kinds of tough choices have to be made by individuals and where the interests of different groups of people are indeed in conflict. This view of work is consistent with some broad assumptions of economists. Indeed, the economist's concept of the 'net advantages' of work suggests that individuals 'calculate' the overall characteristics of a job of work as against other jobs.

The kind of employment with the most obvious set of 'net advantages' is one which provides security of employment, pleasant conditions, high pay, high status, and a secure future 'career' – rounded off at retirement with a secure fat pension. Few occupations, if any, can offer all of these advantages – and those which can offer the closest approximation to the ideal, for the greatest proportion of entrants, are able to select these entrants from a much larger number of aspirants. From this point of view the unskilled and casually employed manual labourer has very many disadvantages; near the opposite pole one example would be the British administrative grade civil servant – who has absolute security of tenure, relatively interesting and important work, and a fairly certain 'career' involving substantial promotion. But in many – perhaps most – occupations the career opportunities are less clear-cut. There is considerable uncertainty. For instance, in many manual occupations there is some chance of becoming foreman; in many white-collar and clerical occupations there is a rather better chance of a managerial post. And so on. But whether the chance is one in 100 or one in 10 or one in two may depend on such imponderables as the future prospects of the occupation or of the employing organisation itself.

The ideal type of 'career' involves the individual in a series of work roles which are arranged in a hierarchy of status. The ideal type of military career is thus one where a man starts as a private soldier and becomes a General or Field Marshal. In this sense most people's 'careers' do not involve their whole working lives. Most careers arrive at a plateau at some stage before they reach the age of 65, and many careers move downwards long before this.

Even though the individual's information about relevant work careers will usually be very imperfect, nevertheless his imperfect perception of the career patterns will be important in his orientation to work – and employers will be aware that his current pay-package

will not be his only concern. Employing organisations try to offer 'good career prospects' at least to certain categories of employees. Employers and employees are also aware that the continued growth of the employing organisation offers improved career prospects to more people and that this has implications for the long-term management of the enterprise – contributing especially to the growth impetus (or 'empire building') which is often noted as a characteristic of large organisations.

For women, there is the problem of continuing two careers – the work career and the home career. The demands upon a woman's energies and emotions at home can be expected to depend largely upon her children – their number and their ages. Many women deliberately choose to emphasise their home career and to forego career advancement at work. For women in 'professional' work – with typically lengthy training, and competition for career advancement – the phasing of home and work careers may pose especially difficult problems.

Many men realise from a very early age that they have reached the limit of their likely career advancement. A team of sociologists who studied car assembly and other affluent manual workers in Luton reported that most workers had an 'instrumental' approach to work based on a realistic view of the poor promotion chances and the physical disadvantages of the work; but the workers saw the fairly high pay as compensating for these disadvantages. The workers tended to focus upon the domestic comforts which their pay-packets could earn.

These authors (Goldthorpe, Lockwood *et al*, 1968, p. 175) predicted that the instrumental orientation of the Luton manual workers they studied would become more common among men lacking high levels of skill or other qualifications:

In the case of those men at least who do not possess skills which are in high demand, there will be mounting inducements to relegate work to the level of merely instrumental activity and to seek employment which offers a high economic return, if only as compensation for its inherent 'disutility'. Moreover, such a tendency is likely to be encouraged by the re-orientation of working class and community life which would now seem to be taking place on a fairly large scale. One major outcome of this is to bring the conjugal family into a more central position than previously in the life of the manual worker, and thus to widen and strengthen the expectations which are held of him as a husband, father and family provider. And to the extent, then, that his out-of-work life becomes dominated by home and family concerns, the link between this and the worker's occupational life is likely to be narrowed down to one of a largely economic kind. In other words, a privatised social life and an instrumental orientation to work may in this way be seen as mutually supportive aspects of a particular life-style.

On the other hand, it is often suggested that a particular group of manual workers or white-collar workers are 'not acting in their own interests' – for instance, by not accepting a productivity deal. This often refers merely to their immediate economic interests. We need, however, to know more about their non-economic interests before we are justified in making such a judgement. The major life interests of some people appear to involve their work; and for others to involve their lives outside work. Moreover, even within one single occupational group there may be different individuals with differing orientations: one man 'living for his work', while another is a 'family man'.

6.2 Leisure and Work

The same man who goes to work also has leisure. And 'leisure' is not necessarily something which takes place away from the office or factory; similarly 'work' is not necessarily something which takes place away from home. Open University students are not the only people who may be studying hard *at home* with a view to furthering their careers *at work*. Those who 'work' at home are not confined to the self-employed. In many white-collar and professional occupations people complain that they can't get any work done at the office – the 'real work' is contained in the bulging briefcase which is taken home.

6.3 The Drive for 'Autonomy' in Work

In much economic behaviour away from work, people strive to further their own *autonomy* – to increase the area in which they are free to take their own decisions. Most groups of workers do manage to exert a good deal of *tactical* control over working conditions. Sometimes this is directly related to economic benefits. D. Roy's article in the *Reader* (pp. 235–45) about piece-workers illustrates some of the many devices workers have for using their tactical autonomy in order to manipulate the 'rate for the job'.

Other studies show that even under a prevailing system of piece rates individual workers usually do not try to *maximise* their own earnings. A norm for each man or each woman for himself or herself may operate in some very individualistic work situations; and in other cases one or two individuals may try to maximise their earnings by spectacular feats of individual output. But more usually a group norm develops of what a standard day's effort is. The exact character of these group norms will depend upon a large number of variables. Moreover the existence of such norms is well known to supervisors and managers, and the findings of early pieces of industrial sociology have long since become incorporated into the textbooks of business schools and into conventional management thinking. The management has often tried to encourage certain sorts of 'informal' group behaviour – just as Generals try to improve front line morale perhaps by turning a blind eye to this and 'unofficially' encouraging that. When 'informal' practices receive 'formal' support it becomes increasingly difficult to distinguish one from the other (and the terms 'formal' and 'informal' are now regarded by many sociologists as terms too vague and misleading for use).

'Informal' norms are, moreover, 'formally' adopted as negotiations between management and trade unions become more sophisticated. 'Rules' of various kinds are agreed. Fair levels of output – related to workers' norms in some way – are written into trade union rule-books. Safety regulations are agreed. In many kinds of work there is also detailed legislation covering conditions of work.

All of these codified rules allow for further action from workers directed towards maintaining or increasing tactical autonomy. Regulations and rules are designed for general application across a whole industry or occupation and may thus be imprecisely formulated. Such rules are usually in part either contradictory, superfluous, or irrelevant in the context of any particular job. Thus the rules allow the worker who so wishes to engage in various practices which will increase his control over his immediate working situation. There are many variations on this theme. One of the best-known collective

versions is the 'work to rule'; here the workers merely undertake to observe strictly all the relevant rules.

6.4 The Goals of Occupations

Do the members of an occupation try to increase their earnings? On the whole the answer is probably 'yes', but there are many different sorts of occupations and different ways of setting about such a goal. What is an occupation? There can be no simple definition – occupations are always changing, splitting up, merging, disappearing and emerging. Each time some major new technology appears – for instance, aircraft or computers – a new range of occupations develop. The 'What is an occupation?' question presents problems for trade unions which set out to further the interests of the members of an occupation. Do these trade unions seek to maximise the earnings of their members? This is usually one of their stated goals. But a large trade union is itself a complex organisation with a certain internal dynamic of its own. Some membership groups within a trade union may be better than others at exerting pressure on the leadership. Moreover any specific trade union operates to a great extent in relation to the main employers of its members – and thus its goals will depend partly upon what goals those employing organisations pursue.

Some occupations use strategies designed to out-manoeuvre rival occupations. These struggles may take place within a single trade union or within a single employing organisation. Frequent battlegrounds in such struggles are new areas of work, new types of skills, new contracts. In the course of these strategies such goals as maximising the earnings of current members may be sacrificed (or alternatively given extra prominence).

A strategy common to many trade unions and to 'professional' associations is the attempt to restrict entry to the occupation. Professions, of course, have a number of explicitly non-economic goals. By developing a body of theoretical knowledge, 'rational' qualifications and training, some other occupations attempt to move along the various dimensions of 'professionalisation'. The professions are of great interest to sociologists – especially in view of recent developments which place many professionals inside large organisations, such as hospitals and commercial companies.

One of the key characteristics of an established profession such as law or medicine is that it is given a legal right to control entry and to license practitioners. Critics of the professions claim that this is a legal monopoly. Whether prestige produces high pay or high pay produces prestige is a question of more than academic interest. There seems to be some evidence for the suggestion that high prestige comes first and leads to high pay. If this is so, it presents many dilemmas for such 'semi-professions' as nursing, social work, schoolteaching, journalism and the like. Often these semi-professions are torn between a trade union and a professional strategy towards increasing the prestige and the earnings of occupational members. Members of such occupations typically are also subject to rapid mobility between jobs, major internal differences in perceived status, and considerable geographical segmentation.

In these and other occupations it may be unrealistic to regard the whole occupation as having any single predominant goal – of an economic, or any other nature. An occupation is perhaps most accurately seen as having a collection of goals, some economic and

some non-economic, some long-term and some short-term, some pursued by a majority and some by a minority of members. These goals change with time as a result of shifting coalitions. Often the type of coalition goal for which wide support can most easily be mobilised within the occupation will be an economic goal – such as maintaining or increasing earnings, relative to other specific groups; but this coalition goal will usually be the product of other ‘non-economic’ social processes, values and goals.

7 CAPITAL AND ENTERPRISE: THE GOALS, OWNERSHIP AND CONTROL OF COMMERCIAL ORGANISATIONS

7.1 Business Goals

Do business organisations have purely economic goals? Do they sacrifice all in a remorseless pursuit of profit maximisation? Obviously the real world will be somewhat more complicated than such questions imply. Obviously also there are several different possible sorts of economic goal for a conventional western public company, such as:

- 1 Maximising the dividends distributed to shareholders.
- 2 Maximising capital investment.
- 3 Maximising sales volume.
- 4 Maintaining stability or maintaining a particular rate of growth, in one or more of the first three.

The idea of *maximising* any one kind of economic goal may be rather improbable. Ruthless pursuit of any kind of economic goal may prove counter-productive – it may needlessly antagonise important interests, such as the customers, the competition, or the government. The government in almost any type of political system is of great importance to business organisations. A government may have a special taxation policy – for instance, in the case of foreign oil companies; or it may have a special subsidy policy, for instance in the case of building new factories in designated areas. Business organisations obviously take such points into account. But ‘prestige’ of a general kind is also a very important business asset.

We have already noted some of the goals of occupations – and any large business organisation must devote a great deal of attention to its various categories of employees; some of the occupational wants will be of a directly economic kind, others much less obviously so. The goals of a business organisation will also depend partly upon its customers – whether the customer is the government, or other companies, or a limited sector of the public, and so on. Other obvious categories of people with some influence over the goals of the organisation are the shareholders and the senior management. This raises the famous conundrum of ownership and control.

7.2 Ownership and Control

Are traditional economists’ terms for such factors of production as Capital and Entrepreneurship any longer valid? Economists themselves have often asked such questions. For instance, the great English economist, Alfred Marshall, in his book *Industry and Trade* (published 1919) discussed the question of the divorce of ownership and control.

The *managerialist* argument takes many forms, but all are based

primarily upon the broad spread of the shareholdings of a modern public company. The shareholders – it is argued – although nominally owning the company have no way of mobilising their potential control; the company in practice is controlled by professional managers – who although they own a negligible proportion of the shares still make all the basic decisions. Moreover, these controlling managers may well make decisions which are in line with their own managerial interests rather than with those of the shareholders. Some people have suggested, for instance, that the manager-controllers of a public company will tend to play down the importance of distributing dividends to shareholders and will give more emphasis to ‘public responsibility’ and other prestige-conferring goals.

The counter-arguments also take many forms. Some people point out that although only a few large companies still have a dominant shareholder or controlling family, nevertheless there are often a few powerful institutional shareholders – such as large insurance companies or other large public companies. Another counter-argument is to the effect that managers – despite not having large personal shareholdings – will tend to align themselves with the major ownership element. One subtle version contends that although the managers may personally own a tiny fraction of the total shares, nevertheless individual managers will often own blocks of shares which are of major significance to their personal incomes.

A number of empirical studies of various aspects of this problem have been undertaken in the U.S.A., Britain and other countries. Some studies focus upon the relative importance of large companies in the economy, others upon the shareholding structure of particular large companies, and yet other studies seek to relate such variables as the distribution of shareholders’ dividends to such structural variables as the composition of the board of directors.

Most of the studies appear to report evidence upon the *potential* for control (for instance, that a certain number of insurance companies own a certain proportion of shares); there are few studies which show how such potential control is actually executed or how specific decisions are made. These studies, moreover, suggest that the dichotomy of either ownership control or of managerial control is false. The answers are confused partly because the question posed at the start of this section is also confused. One recent writer (Child, 1969, p. 51) on the topic reports:

In short, the popular managerialist thesis that the separation of business ownership and control is one of the most momentous developments in modern capitalist society fails to remain convincing when placed alongside the available evidence. This evidence suggests instead that in such a society there are wider economic, social, and technological constraints operating on the business enterprise which tend to minimise the behavioural differences between owner-managers and non-owner managers. This is not to say that the *alternative Marxist thesis* can therefore be accepted uncritically. This too makes some over-simplified assumptions regarding, for example, the range of objectives held by owner-managers or the deterministic importance of property ownership *per se*.

Despite such disagreements, false dichotomies, and false assumptions, there is, however, much more widespread agreement as to the great importance of the *social context* and the social and political value system within which economic production is carried on.

8 CONCLUSION: ACHIEVEMENT AND ASCRIPTION IN THE ECONOMY

This unit has paid little attention to the relations between economics and sociology as academic disciplines. But it is difficult to summarise neatly the central ideas with which the 'sociology of economic behaviour' deals, partly because sociologists themselves more often talk about 'industrial sociology', the 'sociology of organisations', the 'sociology of occupations and professions', and 'social stratification'. The tendency of sociology has always been to see the economy within a broader social framework. When Parsons described the economy as one of the sub-systems of society he was only following Max Weber's classic *The Protestant Ethic and the Spirit of Capitalism* (1904). In this study Weber had seen the rise of Protestantism in central Europe in the early sixteenth century as preparing the way for, and providing an ideology suitable to, the rise of capitalism.¹ Weber stressed not only the puritanical aspects of protestantism – which were favourable to the accumulation of capital – but he also stressed the importance of achievement.

The conceptual distinction between *achievement* and *ascription* can be related to most of the content of this unit. One of the key distinctions between the values prevalent in the family and at work lies in the different forms of recruitment. The family recruits its new members by ascription – they are born into it; whereas most people in industrial societies are recruited into work-roles (and follow their work careers) on a basis of achievement. In the *Reader* article on Japanese Labour-Management relations, both kinds of recruitment were illustrated – executive personnel being recruited on the basis of achievement through competitive examinations, whereas manual workers were recruited on the basis of ascription through family ties. The discussion about the underlying social values in the American and British economies also stressed what was claimed to be the greater emphasis on achievement in American industry.

In the case of property, also, these contrasted principles are of paramount importance. One of the most contentious aspects of the various forms of 'property' is that much – but by no means all – property is inherited; much property, therefore, is acquired (for instance, by children from their parents) on the basis of ascription – not achievement.

Of course in most systems there is a mixture of ascription and achievement – and this mixture is very evident in differing orientations to work. The whole concept of 'career' stresses achievement, but getting on to a suitable career ladder in the first place may depend heavily upon ascription – being born into a certain kind of family. Moreover, when a manual worker decides to concentrate on earning a higher wage – emphasising his privatised home life and foregoing ambition – he is in effect saying that given his situation the possibilities for substantial achievement are negligible. Finally, in examining capital and enterprise, or ownership and control, we can once again see the contrasted principles of achievement and ascription – the notion of managerial control (and the importance of ownership) again stresses achievement.

¹ There is no room here to discuss the validity of Weber's thesis.

BIBLIOGRAPHY

- BLAU, PETER (1964). *Exchange and Power in Social Life*. London, Wiley.
- BURRAGE, MICHAEL (1969). 'Culture and British Economic Growth', *British Journal of Sociology*, XX, No. 2. June 1969.
- CHILD, JOHN (1960). *The Business Enterprise in Modern Industrial Society*. London, Collier-Macmillan.
- CHINOY, ELY (1967). *Society*. New York, Random House.
- DURKHEIM, EMILE (1893). *The Division of Labour*. London, Collier-Macmillan.
- GOLDTHORPE, JOHN H., LOCKWOOD, DAVID, *et al.* (1968). *The Affluent Worker: Industrial Attitudes and Behaviour*. Cambridge, Cambridge University Press.
- JOHNSON, HARRY M. (1961). *Sociology: A Systematic Introduction*. London, Routledge.
- MARSHALL, ALFRED (1919). *Industry and Trade*. London, Macmillan.
- MUSGRAVE, PETER (1969). *The Economic Structure*. London, Longmans.
- PARKER, S. R., BROWN, R. K., *et al.* (1967). *The Sociology of Industry*. London, Allen and Unwin.
- PARSONS, TALCOTT, and SMELSER, NEIL J. (1956). *Economy and Society: A Study in the Integration of Economics and Social Theory*. London, Routledge.
- SMELSER, NEIL J. (1963). *The Sociology of Economic Life*. New York, Prentice Hall.
- STONE, G. P. (1954). 'City Shoppers and Urban Identification', *American Journal of Sociology*, LX, No. 1, pp. 36-45.
- WEBER, MAX (1904). *The Protestant Ethic and the Spirit of Capitalism*. London, Allen and Unwin.
- WEBER, MAX (1964). *The Theory of Social and Economic Organization*. New York, Free Press.
- WORSLEY, PETER (1970). *Introduction to Sociology*. London, Penguin.

Notes

1 WHY PEOPLE LIVE IN SOCIETIES

- 1 The Fundamentals of Human Nature
- 2 Men and Government
- 3 The Economic Basis of Society
- 4 Societies and Environments
- 5 Man as a Social Animal

2 HOW PEOPLE LIVE IN SOCIETIES

Socialization

- 6 Child Socialization
- 7 Personality Development
- 8 Attitudes and Prejudice
- 9 The Family and its Functions

Economy and Society

- 10 Economic Wants
- 11 Production and Supply
- 12 Markets and Prices
- 13 The Sociology of Economic Behaviour

Money, Wealth and Class

- 14 The Working of the Economy
- 15 Money
- 16 Distribution of Income
- 17 Social Stratification
- 18 The Psychology of Social Class

Spatial Aspects of Society

- 19 Rural Land Use
- 20 Location of Industry
- 21 Zoning within Cities
- 22 The Size and Spacing of Settlements
- 23 Approaches to Political Geography

Government and Politics

- 24 The Formal Structure of Government
- 25 Informal Political Institutions
- 26 Government and Politics without the State
- 27 (i) Political Culture (ii) Politics and Political Systems
- 28 Politics in Groups

Stability Change and Conflict

- 29 Stability and Function in Society
- 30 Social Change in Society
- 31 Stability and Change in Social Groups

3 WHAT KIND OF PROBLEMS PEOPLE FACE IN SOCIETIES

The Population Explosion - an Interdisciplinary Approach

- 32 The Demographer and his World
- 33 Demographic Regions of the Indian Sub-continent
- 34 Population and Economic Growth
- 35 Population Growth and Social and Political Systems
- 36 Diffusion and Acceptance of Change

D100 – UNDERSTANDING SOCIETY: A FOUNDATION COURSE

1 WHY PEOPLE LIVE IN SOCIETIES

- 1 The Fundamentals of Human Nature
- 2 Men and Government
- 3 The Economic Basis of Society
- 4 Societies and Environments
- 5 Man as a Social Animal

2 HOW PEOPLE LIVE IN SOCIETIES

Socialisation

- 6 Child Socialisation
- 7 Personality Development
- 8 Attitudes and Prejudice
- 9 The Family and its Functions

Economy and Society

- 10 Economic Wants
- 11 Production and Supply
- 12 Markets and Prices
- 13 The Sociology of Economic Behaviour

Money, Wealth and Class

- 14 The Workings of the Economy
- 15 Money
- 16 Distribution of Incomes
- 17 Social Stratification
- 18 The Psychology of Social Class

Spatial Aspects of Society

- 19 Rural Land Use
- 20 Location of Industry
- 21 Zoning within Cities
- 22 The Size and Spacing of Settlements
- 23 Approaches to Political Geography

Government and Politics

- 24 The Formal Structure of Government
- 25 Informal Political Institutions
- 26 Government and Politics without the State
- 27 (1) Political Culture (2) Politics and Political Systems
- 28 Politics in Groups

Stability Change and Conflict

- 29 Stability and Function in Society
- 30 Social Change in Society
- 31 Stability and Change in Social Groups

3 WHAT KIND OF PROBLEMS PEOPLE FACE IN SOCIETIES

The Population Explosion – an Interdisciplinary Approach

- 32 The Demographer and his World
- 33 Demographic Regions of the Indian Sub-continent
- 34 Population and Economic Growth
- 35 Population Growth and Social and Political Systems
- 36 Diffusion and Acceptance of Change

